NASA Contractor Report 172118—Vol-2

National Aeronautics and Space Administration (NASA)/
American Society for Engineering Education (ASEE)
Summer Faculty Fellowship Program--1988

Volume 2

Richard B. Bannerot, Editor
University of Houston - University Park
Houston, Texas

Stanley H. Goldstein, Editor
University Programs Office
Lyndon·B. Johnson Space Center
Houston, Texas

NASA

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas

PREFACE

The 1988 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA) / American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began in 1965 at JSC and in 1964 nationally, are

1. To further the professional knowledge of qualified engineering and science faculty members

2. To stimulate an exchange of ideas between participants and NASA

3. To enrich and refresh the research and teaching activities of participant's institutions

4. To contribute to the research objectives of the NASA Centers

Each faculty fellow spent at least 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA / JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1988. Volume 1 contains reports 1 through 14, and volume 2 contains reports 15 through 26.

CONTENTS

Volume 1

OPTICAL JOINT CORRELATION USING THE
DEFORMABLE MIRROR DEVICE

Final Report

NASA/ASEE Summer Faculty Fellowship Program

Johnson Space Center

Prepared by: Jerome Knopp

Academic Rank: Associate Professor

University and Department: University of Missouri-
                                      Columbia/Kansas City
                                      Department of Electrical
                                      and Computer Engineering
                                      Truman Campus
                                      600 West Mechanic
                                      Independence, MO 64050

ABSTRACT

An experimental investigation of the Deformable Mirror
Device (DMD) developed by Texas Instruments at Dallas for
use in object identification was completed. The DMD was
tested as a joint correlator. The DMD was used as a spatial
light modulator on which the squared modulus of the Fourier
transform of test object pairs was written. The squared
modulus was phase encoded on the DMD after it had been
thresholded and rewritten as a binary phase function. The
thresholding was found to produce a sharp peak in the
autocorrelation when the test objects were matched and no
significant peak in the case of distinct objects. It was
concluded that the use of the DMD as a joint correlator
looks promising and further studies should be carried out.

INTRODUCTION

The location and recognition of objects belonging to small well defined sets (e.g., handtools drifting in space) or objects belonging to large less well defined classes (e.g., boulders on the surface of Mars.) are typical parts of mission requirements on several NASA programs including the planned unmanned expedition to Mars, the EVA (Extra Vehicular Activity) retriever program and autonomous rendezvous and docking. In most cases the speed of the recognition process must rival that of the human eye-brain. This is difficult competition. Even the fastest available digital computers would not be able to process data fast enough to keep pace with human vision. An interesting alternative to digital computing is optical computing. Optical computing takes advantage of massive parallel processing. Within the last 10 years, optical processing techniques have matured to the point that they are worthy of consideration in tasks where speed is a premium. Although it lacks the flexibility of digital processing at the present, it can perform certain fixed tasks, such as Fourier transforms in two dimensions, almost instantaneously (i.e., essentially in the length of time it takes the light to pass through the optical processing system.) At present, the Tracking and Communications group at NASA is studying the use of a hybrid system that exploits the best of both digital and optical processing for artificial vision problems. The key element in this approach is the Deformable Mirror Device (DMD), a new type of integrated optics device under development by Texas Instruments (TI) in Dallas.

The DMD consists of a 1/4" square array containing 128 x 128 mirror elements. Each mirror element consists of a set of four individual cantilevered mirrors arranged as in a 2 X 2 matrix (roughly in the form of a cloverleaf.) Each of the mirror elements can be deflected as a group using the electrostatic forces determined by capacitively stored charge between a mirror and the array back plane. The charge stored, and thus the deflection, is determined by on-chip electronics and addressing. The DMD is, in essence, a deformable mirror surface whose shape can be digitally controlled. The mirror is also fast. Each element can be deflected at an 8 kHz. rate, although the current experimental setup operates the DMD at 50 Hz. to maintain compatibility with a 50 Hz. European standard video.

At the present, the DMD is being studied for use as an optical correlator. Correlation strength measures can be a key feature in vision identification tasks [2]. Correlation, however, has not been normally used in

practical vision systems since it is computationally
intensive and slow. This is not the case using the DMD where
correlation using optical processing is attractive. The two
classical approaches to correlation using optical processing
are Vander Lugt Filtering (VLF) and Joint Correlation (JC).
The use of VLF has been one of the prime thrusts of previous
research on the DMD and is described elsewhere [3]. The JC
approach was developed and studied this summer as a part of
research conducted under the ASEE summer faculty program.
The key results of this work are described here.

The basic approach to JC involves two Fourier transform
operations and one square law (intensity detection)
operation. First two objects to be correlated are optically
Fourier transformed in the same operation. This is done by
placing the objects side by side in front of a lens and
transilluminating them with coherent light. The lens
produces a joint Fourier transform of the objects in terms
of the electric field [5]. The field is then detected with a
square law detector. The squaring operation produces
crossterms containing the Fourier transform of the joint
correlation of the objects. An additional Fourier transform
of the squared field (using another lens) will yield two
crosscorrelation terms. Optically this results in
autocorrelation peaks in the form of intense points of light
when the objects are matched. The degree of correlation as
well as the details of the object structure determine the
"sharpness" of the correlation point. The implementation of
the JC process on the DMD required some modifications of the
traditional JC process. In conventional optical
implementations the square law operation is usually carried
out by detecting the intensity of the joint Fourier
transform on photographic film. A positive transparency of
the transform then serves as an input for the second Fourier
transform operation using a lens that yields the JC. In the
approach used here the film is replaced by the DMD. The DMD
is not primarily an amplitude modulating device; it is for
the most part, a phase modulating device. In this case, the
intensity was phase encoded.

In previous research, using a computer simulation [5],
it was shown that directly writing the intensity as a phase
would yield good quality correlations. This was attempted
initially by writing the squared Fourier transform patterns,
calculated by a computer, directly on the DMD. The DMD was
then illuminated with collimated light from a helium neon
laser. The reflected light was then Fourier transformed with
a 50 cm. lens. The resulting inverse transform was then
imaged on a CCD array camera and observed on a monitor.
These first attempts were negative and no correlation peaks

were observed. The process was reviewed for flaws. One of the major problems in the previous computer simulation is the assumption that the DMD could be accurately modeled using simple reflecting surfaces whose mechanical deflection was directly proportional to the charge pattern written into the DMD. In the real device, this does not appear to be the case. Based on data obtained by TI [6], it appears that there is a nonlinear relation between charge patterns written on the DMD and the resulting pattern of mechanical deflection. Therefore, an alternative approach was tried that appeared to be much more robust.

The key identification feature in the NASA tasks is a sharp autocorrelation peak that shows strong correlation between an object to be identified and a reference object. A sharp peak is due primarily to a strong single frequency sinusoidal component occurring in the joint Fourier transform. The strength of this single frequency component is proportional to the number of equally spaced point pairs on the objects being correlated. If this sinusoidal component can be emphasized in the case of a good match, then the correlation spot will be bright. A simple approach that appears to accomplish this, and completely avoids the nonlinearity of the DMD, is to threshold the modulus squared of the Fourier transform and rewrite the squared transform as a binary object. If the sinusoidal component is present, it will be emphasized as the fundamental component of a square grating. Typical results using this approach are described next.

FINDINGS

The use of JC on the DMD was tried using thresholding. The Fourier transform of test object pairs were calculated, squared and then thresholded by replacing intensities above a constant value (the threshold) by a 1 and intensities below the threshold by 0. In the cases studied, the threshold was set at one percent of the peak value. The resulting pattern was written on the DMD. Then an inverse Fourier transform was carried out optically and the resulting "correlations" were recorded. Quotes are used here to indicate that correlation is used loosely since a nonlinear mapping has replaced the joint Fourier transform intensity with a binary object.

Figs. 1 and 2 show two sets of objects that were used to test the DMD as a recognition correlator. Fig. 1 shows the pair SS and Fig. 2 shows the pair ST. In this case, the S and T can be considered as reference objects that are being compared with a test object, i.e., another S. A sharp

Fig. 1. The test pair SS.
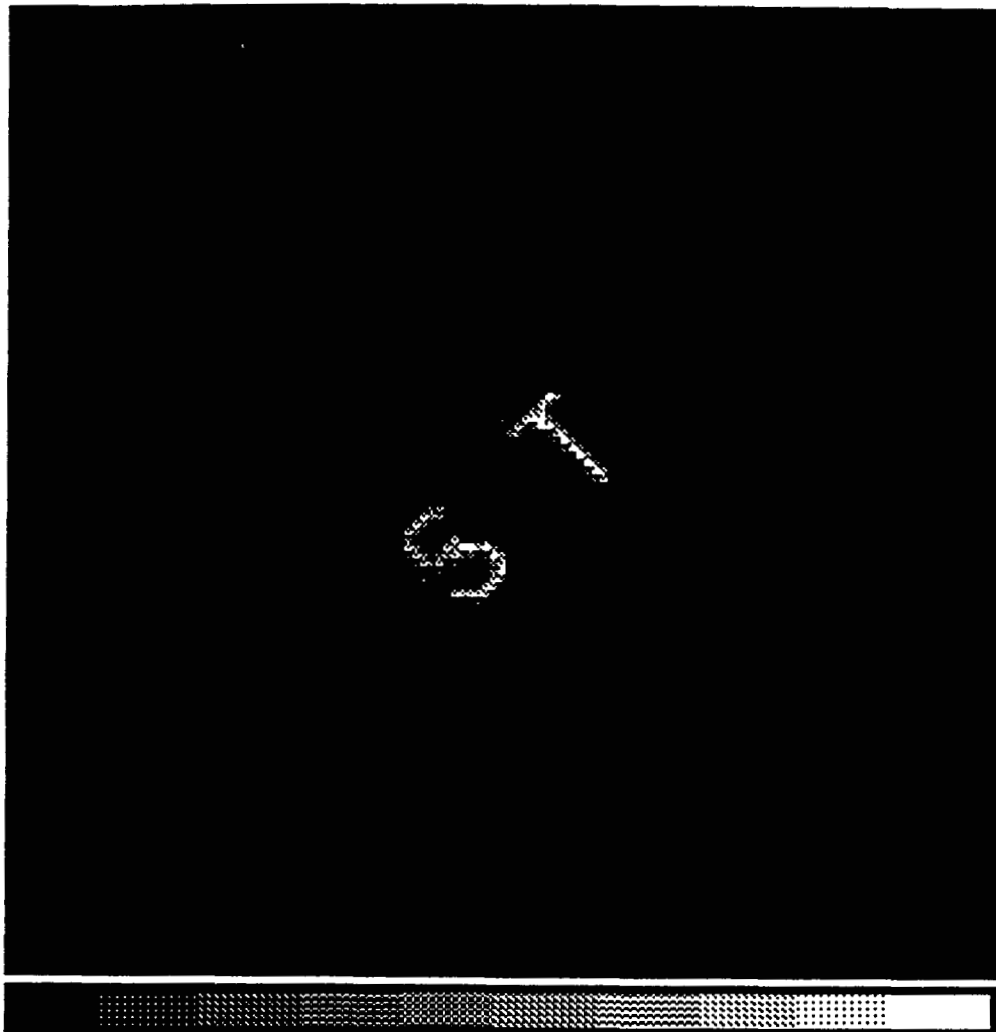
Fig. 2. The test pair ST.

correlation peak is desired when the objects are matched to indicate recognition. However when the objects are not matched then a weaker peak (or a peak too weak to observe) is desired.

Figs. 3 and 4 show the computer generated Fourier transforms after threshholding and converting to binary images for the SS and ST pairs, respectively. Note the strong single spatial frequency component for the SS transform in the diagonal direction. This frequency corresponds to the separation between the objects and will determine where the correlation peak will appear in the output plane. Figs. 5 and 6 show respectively, the resulting correlations made using the DMD. The images were taken from a CCD array camera and stored using a frame grabber. Note the presence of strong off-axis correlation peaks in the case of a match as shown in Fig. 5 and the lack of distinct peaks in Fig. 6. It should be pointed out that all the figures shown are only crudely grey scaled using a 10 level grey scale plotter in conjunction with a laser printer. The grey scale code is shown below the figures.

CONCLUSIONS

The results presented suggest that the DMD can be used to correlate images for identification with good results provided that proper thresholding is used. Since only a small set of objects was tried it is not possible to offer a general scheme for thresholding. It is not known how well the particular threshold scheme used here will work in general. Furthermore, more testing is needed to evaluate scaling, rotation and noise effects using a more interesting object set.

Also, the similarity between the thresholding approach used here and in optical neural networks has not gone unnoticed. Indeed, the only element missing to convert this system to an optical neural net system is a feedback mechanism. This is being considered in future studies using a system the author is presently developing [7].

ADDITIONAL COMMENTS

During the preparation of this final report, but after the author's ASEE summer epoch, some further results were obtained on the JC during a joint "overnight research adventure at NASA JSC " involving Richard Juday and a summer intern student from The University of New Hampshire ,Joe Bailey. The DMD was tested using photographic transparencies as the correlator input. The squared Fourier
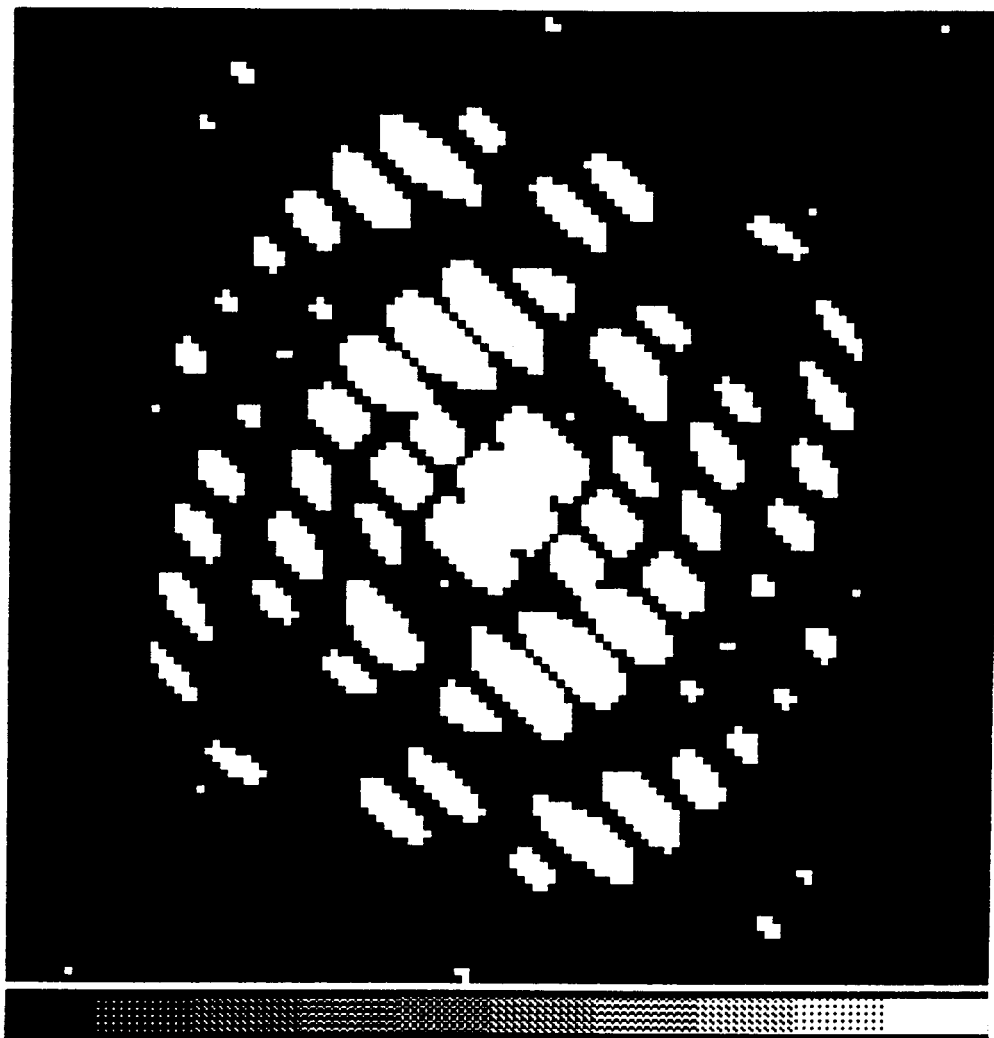
Fig. 3. The computer generated Fourier
transform of SS after thresholding.
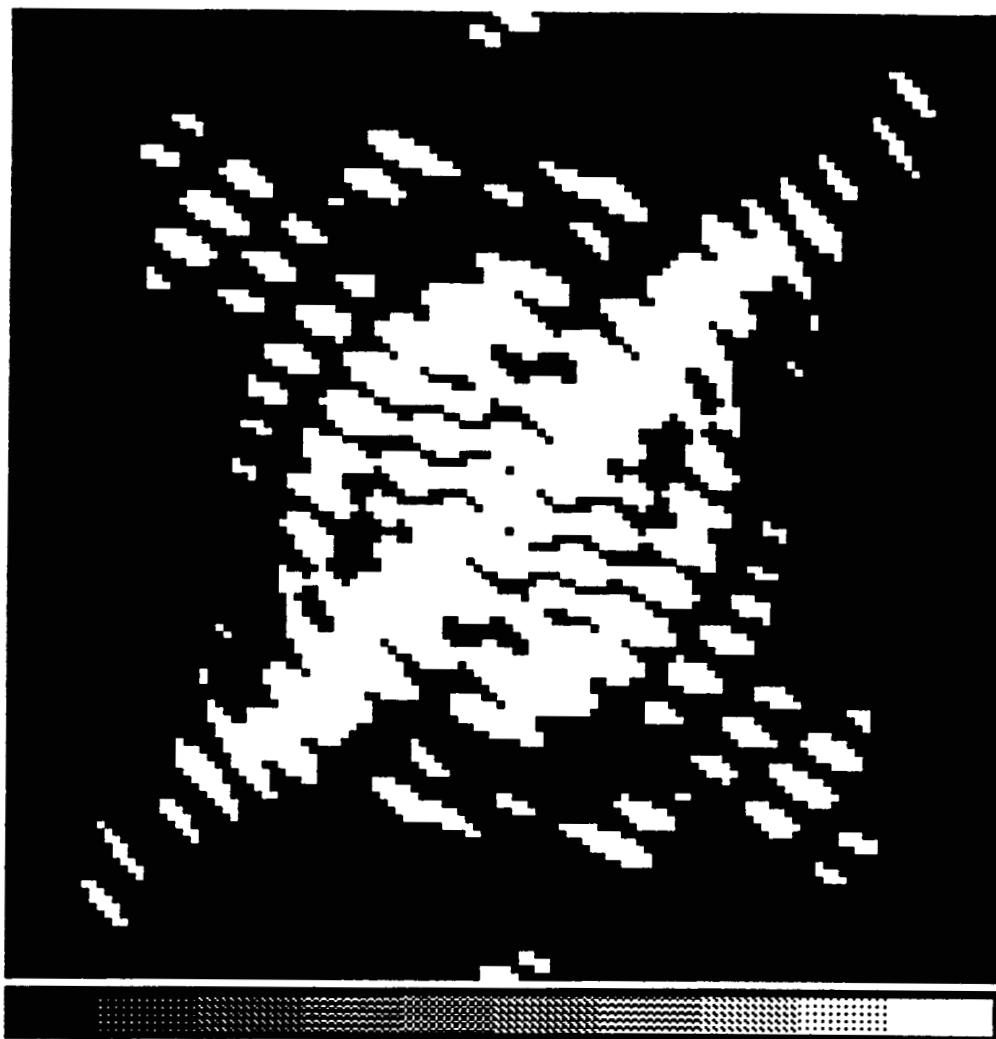
Fig. 4. The computer generated Fourier
transform of ST after thresholding.
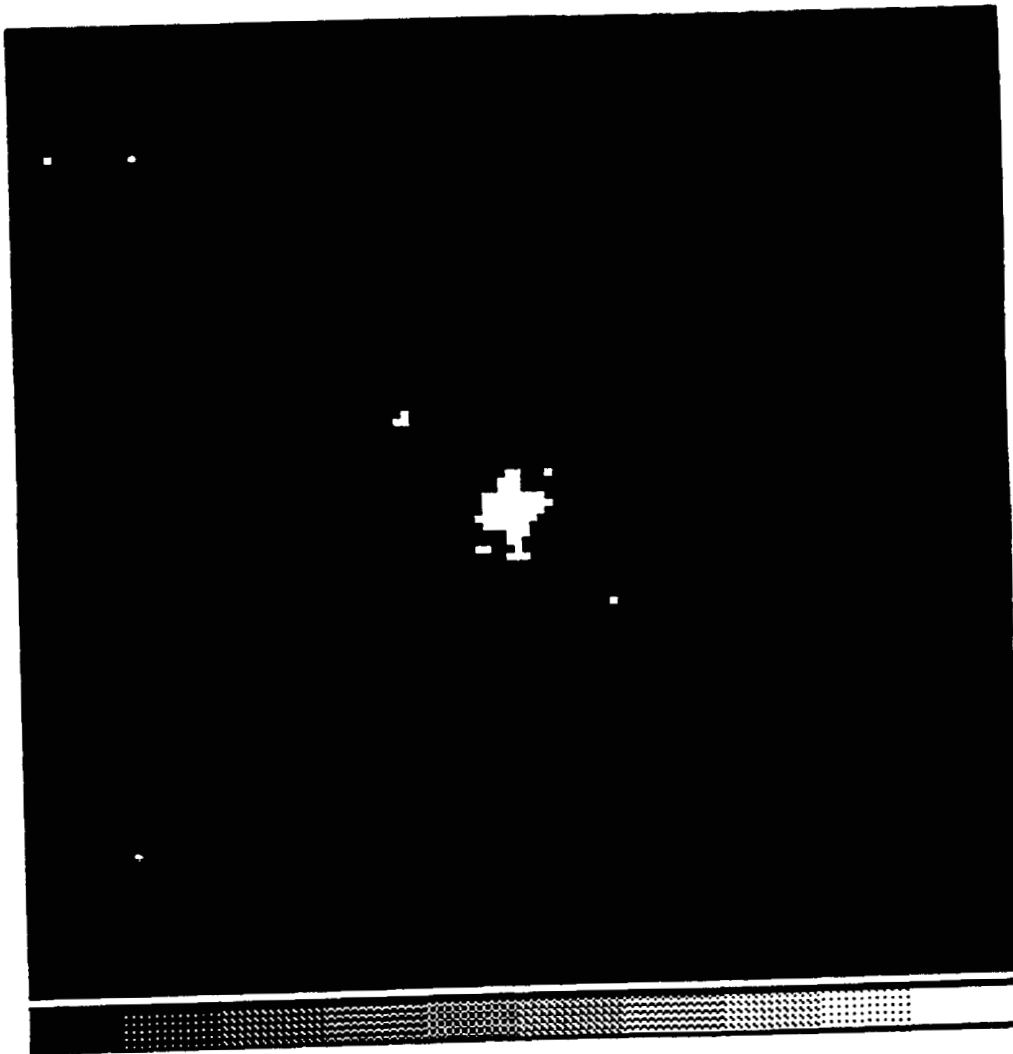
Fig. 5. The joint correlation of the
SS pair.
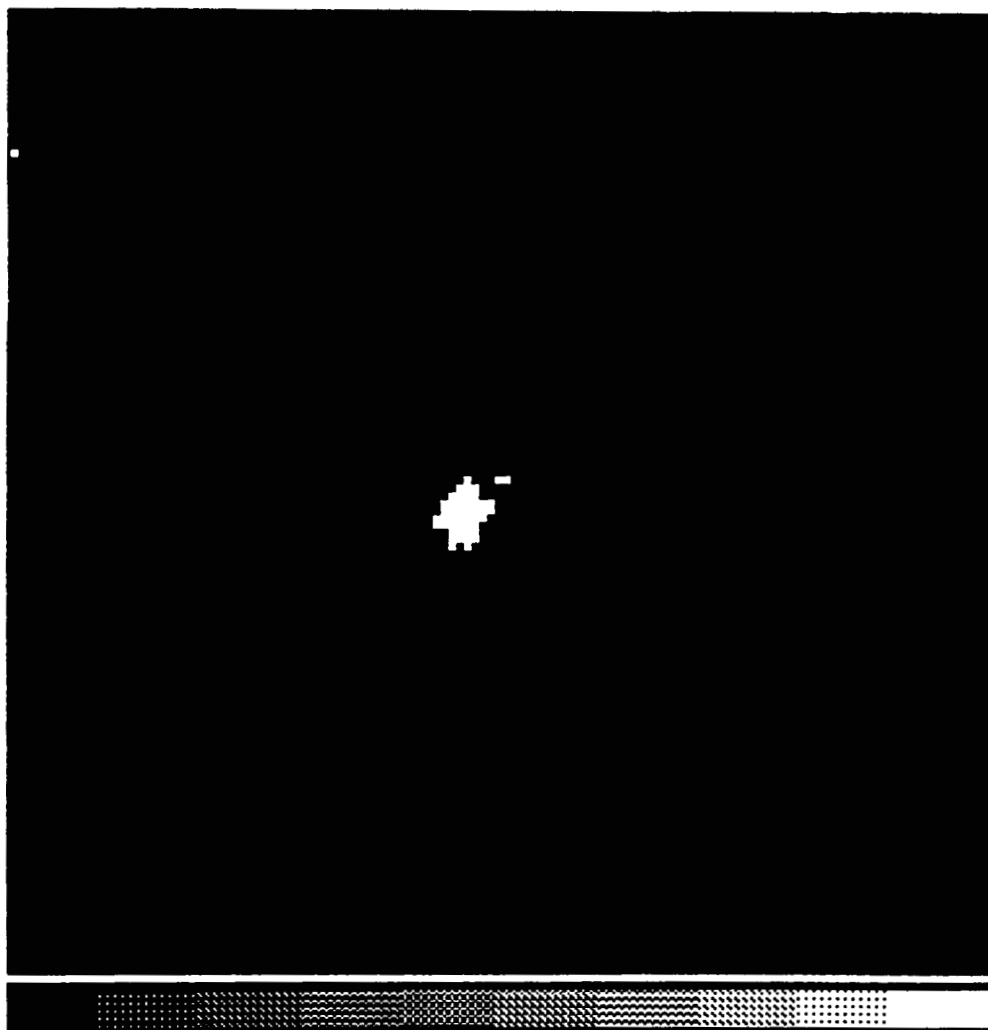
Fig. 6. The joint correlation of the
pair ST.

transform was obtained optically using a lens and directly recorded on a CCD camera. The camera image was written into the DMD using a frame grabber. Good quality correlations were obtained. These results will be reported in detail at a later date. It would appear that real-time JC is possible provided a fast spatial light modulator (SLM) is used as an input to the DMD correlator. Such a real-time device is in the process of being installed. The only additional requirement is a fast frame grabbing routine for transferring the Fourier transform intensity to the DMD. This should not be a problem at the 50 Hz. frame rate now being used to write to the DMD.

REFERENCES

(1) Wu, W.R., R.O. Gale, L.J. Hornbeck and J.B. Sampsell, "Performance of an Improved Deformable Mirror Device", Proc. SPIE, v24, p823, 1987.

(2) Tow,J.T. and R.G. Gonzales, Pattern Recognition Principles, Addison-Wesley, Reading MA, 1974.

(3) Juday,R.D. and B.J. Daiuto, "Relaxation Method of Compensation in an Optical Correlator", Opt. Eng., v26, p1094, 1987.

(4) Goodman J.W., Introduction to Fourier Optics, McGraw-Hill Book Company, New York, 1969.

(5) Giles,M.K., J.M. Taylor, N. Grijalva and B. Gioannini, "Using Deformable Mirror Devices in Correlation", Research report for Texas Instruments, Dept. of Electrical and Computer Engineering, New Mexico State University, Las Cruces NM, Nov. 1986.

(6) Personal communication with Jim Florence at Texas Instruments in Dallas.

(7) Pottinger,H.J. and J. Knopp, "An Optical Neural Network Using Pseudo-Conjugation and Digital Feedback", 31st Annual Midwest Symposium on Circuits and Systems", St. Louis MO, Aug. 1988.

INTELLIGENT CONTROL OF ROBOTIC ARM/HAND SYSTEMS FOR THE
NASA EVA RETRIEVER USING NEURAL NETWORKS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared by: Robert A. McLauchlan, Ph.D.,P.E.

Academic Rank: Associate Professor

University & Department: Texas A & I University
                                  Civil & Mechanical Engineering
                                  Department
                                  Kingsville, Texas 78363

NASA/JSC

Directorate: Engineering

Division: Crew & Thermal Systems

Branch: Special Projects (EC5)

JSC Colleague: Cliff Hess

Date Submitted: September 5, 1988

Contract Number: NGT 44-005-803

ABSTRACT

Adaptive/general learning algorithms using varying neural network
models are considered for the intelligent control of robotic arm
plus dextrous hand/manipulator systems. Results are summarized and
discussed for the use of the Barto/Sutton/Anderson neuronlike, un-
supervised learning controller as applied to the stabilization of an
inverted pendulum on a cart system. Recommendations are made for the
application of the controller and a kinematic analysis for trajec-
tory planning to simple object retrieval (chase/approach and cap-
ture/grasp) scenarios in two dimensions.

INTRODUCTION

Overview

The research work reported herein is important to the future
development of the NASA/JSC EVA Retriever. This highly autonomous,
free-flying robot or robotic system is comprised of MMU, arm and
smart hands. It is being developed to aid crewmen in the perfor-
mance of EVA tasks including the chase, capture and return capabili-
ty required for adrift crewmen or station equipment. The ultimate
goal of the work in developing this system is to enhance the effec-
tiveness of EVA crewmen [1, 23].

The intelligent control of robotic arm/hand systems using
neural network learning controllers is very relevant to EVA Retrie-
ver development. This follows because of the need for autonomous,
adaptive behavior in both planned and unplanned contexts in the
space environment. Neural networks and related advanced learning
controllers offer such capabilities [23].

The work reported herein is concerned with the investigation
and development of neural networks or other types of advanced
learning controllers as:

(a) Supervised controllers with training which because of
their connective, associative memory structure can
develop significant controller generalization capabili-
ty. Such generalization can lead to similar performance
of the retriever arm/hand controller in different but
analogous physical system situations and in stochastical-
ly related loading/excitation environments.

(b) Unsupervised controllers which can self train/adapt to
new learning situations and also exhibit significant
generalization capability. As learning develops, and un-
familiar situations become familiar ones, these neural
networks should provide feedforward compensation with
less compensation via the feedback path [7, 11, 15, 26].

Neural Networks for Intelligent Control

Neural networks are massively parallel, distributed processing systems. They have the ability to continuously improve their performance via dynamic learning [7, 9, 15-18, 36]. As used in this report, neural networks refers to "artificial", i.e., programmable systems of processing elements. As such they form a research area of intense interest in artificial intelligence.

Initial neural network research concentrated on the computationally intensive areas of adaptive signal processing, as, e.g., pattern recognition, real-time speech recognition and image interpretation. Recently there has been a resurgence of interest in neural networks because of (a) Advances in training algorithms for networks, and (b) Availability of extremely fast, relatively inexpensive computers for implementing these algorithms. These developments have lead to the consideration of neural networks for the real-time identification and control of large flexible/articulated aerospace and robotic systems [7, 27, 28].

Neural networks can provide mechanisms for (a) Associative memory, (b) Pattern recognition, and (c) Abstraction. These are emergent properties of networks of neuronlike units with adaptive synaptic connections [10, 14, 22, 29, 32]. These mechanisms arise from the neural network being a system of interconnected "neuronlike" elements modeled after the human brain. This system operates on input data in an "all at once" mode rather than in a conventional computer's "step by step" algorithmic approach [7, 9, 29]. Different learning architectures can be used in training for intelligent control. This is done to provide appropriate inputs to the system so that the desired responses are obtained. Uncertainty and noise can be handled by a neural network via the Hebbian type of associative learning arising from adaptively modified connection strengths [21, 29]. Kawato el al [15-18] indicate that a neural network model can be used to control voluntary movement with applications to robotics. Implemented as a multilayered, hierarchically intelligent control system, neural networks can be implemented to effect the following:

   (a) Pattern recognition/ condition matching
   (b) Trajectory and approach, grasping, etc. opertation
   (c) "Point of view" transformations - as, e.g., visual to
       sensor/end effector to object, etc.
   (d) System (robot, object, etc.) state observer or model
       synthesis and simulation behavior
   (e) Generation of motion/actuator commands.

Adaptive control is useful for systems which perform over the large ranges of uncertainties which result from large variations in physical and operating parameter values, environmental conditions, and signal inputs. However, adaptive control as such (i.e., without unsupervised learning/unanticipated problem solving features) has difficulty with the following generic problems in designing controllers:

   * Sensor data overload - arising from (a) Data redundancy

16-3

per se, and (b) Specialized, rarely required data
* Multi-spectral, multi-sensor data fusion and mapping/use
  in the proper feedback control law
* Need for system robustness to handle large parameter
  excursions
* Required high-speed, real-time control degradation resul-
  ting from time consuming artificial intelligence calcula-
  tions
* Unsolved sensor choice and placement problems for
  robotic/large control systems.

It should be noted that human intervention is used in traditional
control systems operating with large uncertainty. Such interven-
tion is unacceptable in many real-time applications. This is espec-
ially true for the hostile space environment in which the NASA EVA
Retriever is to operate [1, 23]. It means that automatic techniques
for handling uncertainty must be developed. Neural networks show
great promise for the intelligent, unsupervised control of the mul-
tiple arm plus dextrous robotic hands of the Retriever. The next
section of the report describes the author's research work with the
Barto et al intelligent controller which is a special kind of neural
network with associative search and associative critic neuronlike
elements.

ACE/ASE NEURONLIKE LEARNING CONTROLLER

The Barto/Sutton/Anderson adaptive learning controller is
composed of two types of neuronlike elements with significant
unsupervised problem-solving capacities. These elements are the
<u>associative search element</u> (ASE) and <u>adaptive critic element</u> (ACE).
Barto et al 1983 used a single element of each type. Their ASE
element exhibits a learning strategy which is similar to the earlier
"BOXES" adaptive problem solving system of Michie and Chambers
[24, 25]. The ASE/ACE elements embody refinements discussed in the
literature by Barto and colleagues [2-6, 30-31]. They evolved from
the heterostatic brain function and adaptive systems work of Klopf
[19, 20]. Adding a single ACE element improves the learning perfor-
mance over that of a single ASE alone. This can be clearly shown by
comparing the problem-solving capabilities of BOXES with those of a
single ASE/single ACE learning system and solving the control prob-
lem of balancing an inverted pendulum on a cart. It is interesting
to note that strong analogies exist between the behavorial inter-
pretations of the ASE, ACE adaptive elements and animal behavior in
instrumental learning. There are also strong parallels with the
"bootstrap adaption" systems work of Widrow et al [33-35]. This
work considered the (a) punish/reward critical learning and (b) pat-
tern recognizing control problems. Relevant artificial (i.e.,
programmable) neural networks the ASE, ACE neuronlike elements are
significant. This follows because they indicate that if adaptive
elements are to learn effectively as network components, then they
are constrained to have adaptive capabilities at least as robust as
these Barto et al learning controller elements [2].
    Figure 1 depicts the inverted pendulum on a cart system which is
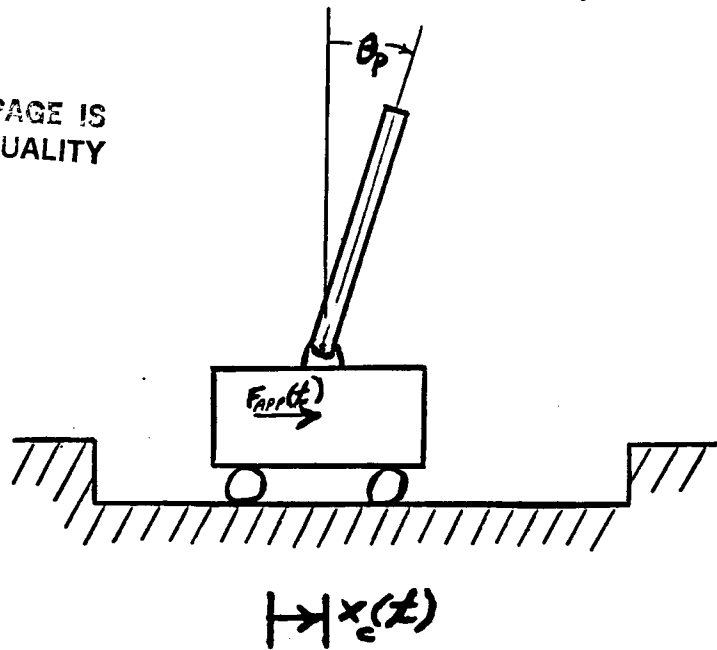
Figure 1.   Representative Model for Cart and Inverted Pendulum
System.
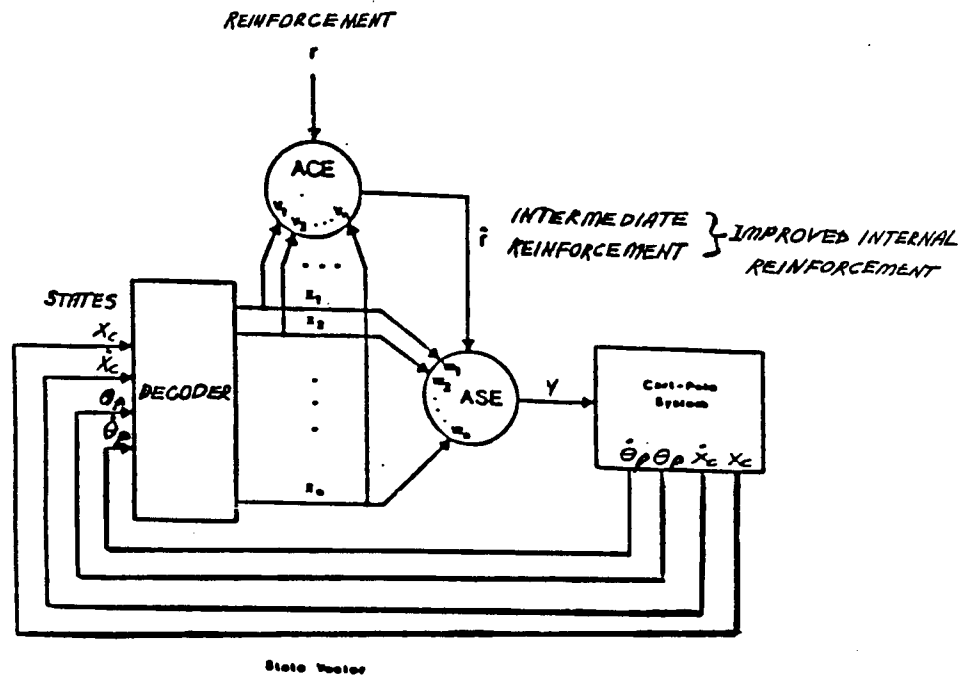
Figure 2.   ASE and ACE Controller for Cart Plus Inverted Pendulum
System

to be controlled. Here the cart can move within the bounds indicated on a one-dimensional track. The pendulum can move only in the vertical plane of the cart and the track. The applied force $F_{app}(t)$ results from the output of the learning controller. It is applied in a bang-bang (+/-) manner and acts with a fixed magnitude to the left or right at discrete time intervals. The pendulum-cart system is described by a four state variable model in the time domain [8]. The four state variables are as follows: (a) X_c - the position of the cart on the track, (b) Θ_p - the angle of the pendulum with the vertical, (c) $\dot{X}_c$ - the cart velocity, and (d) $\dot{\Theta}_p$ - the rate of change of the pendulum angular displacement. The state variable model for this system can be written as

$$
\begin{bmatrix} \dot{X}_c = V_c \\ \dot{\Theta}_p = \omega_p \end{bmatrix}
$$

$$
\begin{bmatrix} \dot{V}_c \\ \dot{\Theta}_p \end{bmatrix} = \begin{bmatrix} [Mc + Mp] & [Mp*L_p*Cos(\Theta_p)] \\ [Mp*L_p^2*cos(\Theta_p)] & [Jp +Mp*L_p^2] \end{bmatrix}^{-1} \begin{bmatrix} [Mp*L_p*sin(\Theta_p)*\omega_p^2 - \mu_c*sgn(V_c) + Fapp(t)] \\ [-Mp*g*L_p*sin(\Theta_p) - \mu_p*\omega_p] \end{bmatrix} \tag{1}
$$

Physical parameters in the above equations specify pendulum length and mass, cart mass, the coefficients of friction between the cart and the track and at the pin connection between the pendulum and the cart, the applied control force, the force due to gravity, and time. Table 1 defines the notation used in equation 1.

The system of first order equations has been solved using second order numerical intergration procedures which have been implemented in the FORTRAN computer program NRLNET. In implementing the learning controller algorithm the state space has been partitioned based on the following 252 quantization interval thresholds:

(1) Cart position X_c: +/- 0.8, +/- 2.4 m ,(4 quantization intervals including failed regions above and below 2.4 m)

(2) Pendulum angular displacement Θ_p: 0,+/- 1, +/- 6, +/- 12 degrees, (7 quantization intervals including failed regions above and below 12 degrees)

(3) Cart velocity $\dot{X}_c$: +/- 0.5, +/-∞ m/s, (3 quantization intervals)

(4) Pendulum angular velocity $\dot{\Theta}_p$: +/- 50, +/- ∞ degrees per second, (3 quantization intervals)

Figure 2 depicts the ASE plus ACE adaptive learning controller of Barto et al [2]. The neuronlike learning system can be described by the following equations:

Element output y(t) which is determined from the decoded state quantization interval vector input

$$
y = f[(\quad w(I,t) * x(I,t)) + n(t)] \tag{2}
$$

16-6

Here the noise $n(t)$ is a real random variable with probability function $p(x)$ and f is either a threshold, sigmoid, or identity transfer function. For the work reported herein, $p(x)$ is the zero mean Gaussian distribution with standard deviation σ, and f is the bang-bang type <u>threshold function</u>:

$$f(x) = \begin{array}{l} +1, \quad x \text{ .EG. } 0 \text{ (applied force action} \\ \quad\quad\quad\quad\quad \text{to the right)} \\ -1, \quad x \text{ .LT. } 0 \text{ (applied force action} \\ \quad\quad\quad\quad\quad \text{to the left)} \end{array} \quad (3)$$

<u>ASE weights $w(I,t)$</u>, 1 .LT. I .LT. N which change over discrete time as follows:

$$w(I,t+1) = w(I,t) + ALPHA * r(t) * e(I,t) \quad (4)$$

In equation 4:

ALPHA   = positive constant determining the rate of change in $w(I,t)$

$r(t)$ = real-valued reinforcement at time t

$e(I,t)$ = eligibility at time t via the input pathway I.

<u>Eligibility traces for the ASE weights</u> which exponentially decay with increasing time, given in equation 5 as:

$$e(I,t+1) = DELTA * e(I,t) + (1-DELTA) * y(t) * x(I,t) \quad (5)$$

in which,

DELTA   = the eligibility decay rate.

<u>ACE weights $v(I,t)$</u>, 1 .LT. I .LT. N which change over discrete time as follows:

$$v(I,t+1) = v(I,t) + BETA * rhat(t) * xbar(I,t) \quad (6)$$

In equation 6,

BETA   = positive constant defining the rate of change of $v(I,t)$

rhat(t)   = $r(t)$ + GAMMA * $p(t)$ - $p(t-1)$, the <u>improved internal reinforcement signal</u> for the critic element

xbar(I,t) = LAMBDA * xbar(I,t) + (1-LAMBDA) * $x(I,t)$, the <u>eligibility traces for the ACE weights</u>

$p(t)$ = $\sum_I v(I,t) * x(I,t)$, the <u>prediction of eventual reinforcement</u>

16-7

GAMMA   = reinforcement learning rate

LAMBDA   = xbar(I,t) trace delay weight

Barto and Sutton [2, 5] explain the derivation of the ACE learning
rule as used above. Additional discussion of the ASE, ACE adaptive
learning controller can be found in references [30, 31].

ASE/ACE LEARNING CONTROLLER RESULTS

This section of the report discusses representative results
obtained by the author with his FORTRAN computer program NRLNET131
implementing the Barto et al ASE/ACE neuronlike learning control-
ler. This program is the result of several modifications by the
author to incorporate general data file input and the file and prin-
ter plot output of the applied control force and the four state
variables as functions of time. The original FORTRAN program
NRLNET00 was the author's implementation of a PASCAL program written
in 1988 by Doug Walker of GHG in support of the Special Projects
Branch (EC5) in the Crew and Thermal Systems Division at NASA/JSC.

TABLE 1. SUMMARY OF PHYSICAL PARAMETER VALUES FOR CART PLUS INVERTED
          PENDULUM SYSTEM

          Mc   = Cart Mass, 1.0 kg
          Mp   = Pendulum Mass, 0.10 kg
          Lp   = Pendulum Length, 0.50 m
          MUc = Cart Coefficient of Coulomb Friction, 0.005
          MUp = Pendulum/Cart Pin Coefficient of Friction, 0.00002
                 N m sec/rad
          Fapp = Magnitude of Force Applied to Cart in x Direction,
                 (+/-) 10 N

TABLE 2. SUMMARY OF THE ASE/ACE NEURONLIKE LEARNING CONTROLLER
          PARAMETERS

          ALPHA = Rate Constant for ASE Weights, 1000.0
          BETA   = Rate Constant for ACE Weights, 0.50
          DELTA = Decay Rate for ASE Eligibility Traces, 0.90
          GAMMA = Learning Rate for Improved Internal Reinforce-
                   ment, 0.95
          LAMBDA = Decay Rate for ACE Eligibility Traces, 0.95
          μ = Mean Value for Gaussian Normal Distribution Used
                   to Define ASE/ACE Output Noise Function, 0.00 and
                   0.10
          σ = Standard Deviation Value for Gaussian Normal Dis-
                   tribution Used to Define ASE/ACE Output Noise
                   Function, 0.01, 0.05, 0.10, 0.15, 0.20, and 0.25

Table 1 gives the physical and control parameter values used in
the simulation work with NRLNET131 for the cart plus inverted pendu-
lum system depicted in Figure 1. Values used for the ASE/ACE neuron-
like learning controller parameters are summarized in Table 2. These
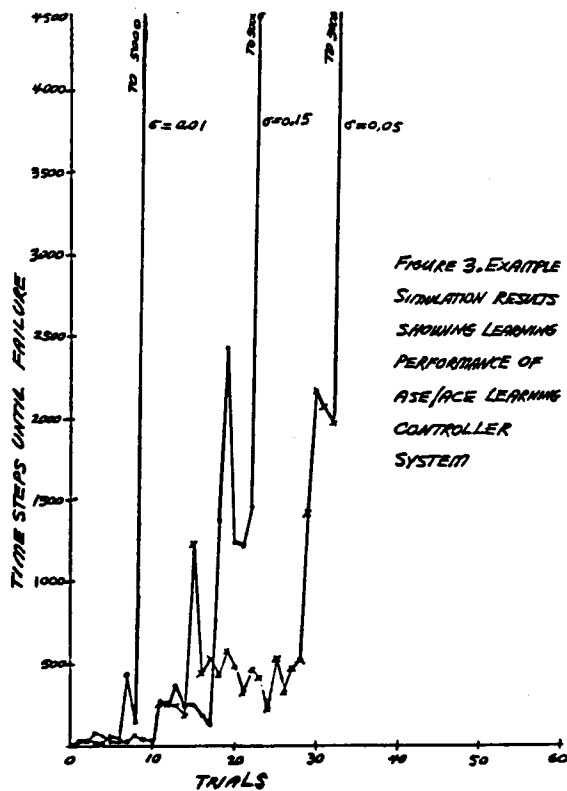
Figure 3. Example Simulation Results Showing Learning Performance of ASE/ACE Learning Controller System
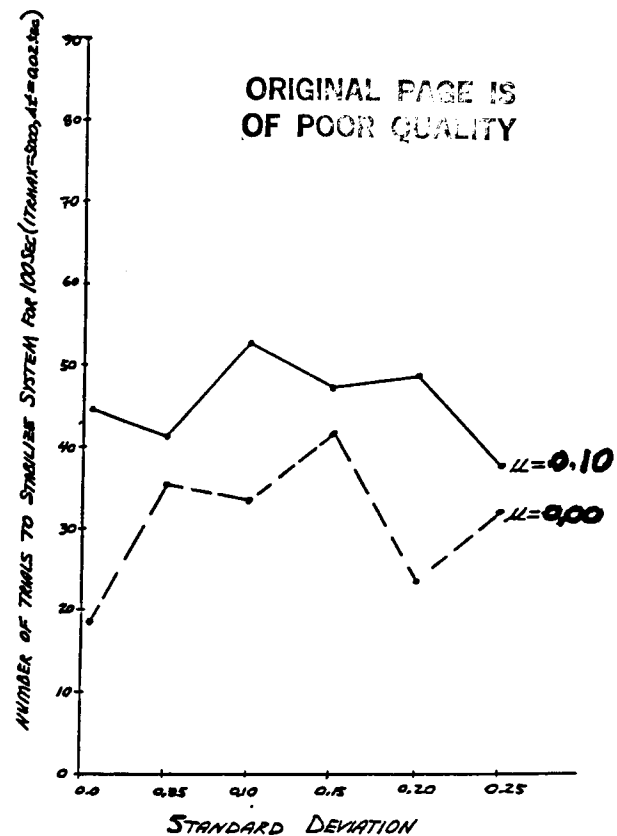
Figure 4. Average Number of Trials for Five Runs as a Function of Standard Deviation With Mean Value as Parameter
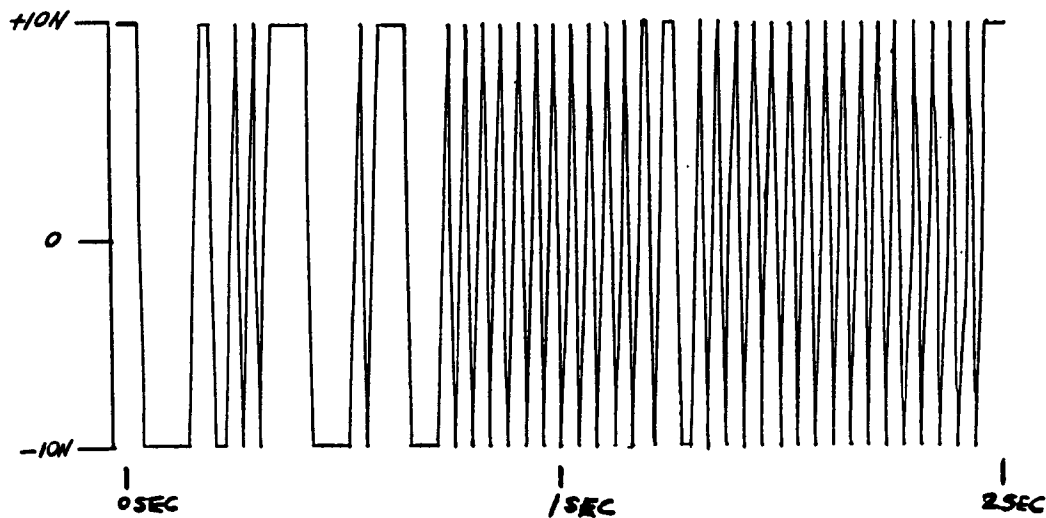
Figure 5. Applied Force Fapp(t), N

Figure 6. Cart Displacement $x_c(t)$, m
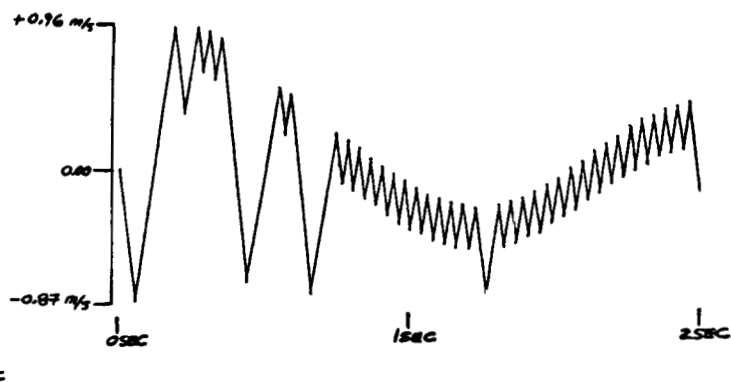
Figure 7. Cart Velocity $\dot{x}_c(t)$, m/sec
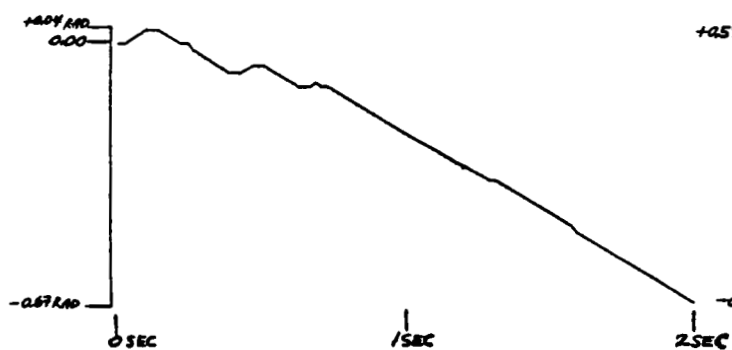
Figure 8. Pendulum Angular Displacement $\theta_p(t)$, rad
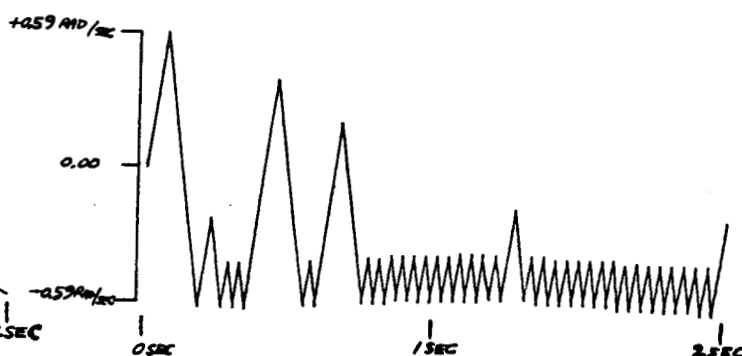
Figure 9. Pendulem Angular Velocity $\dot{\theta}_p(t)$, rad/sec

16-10

parameter values were used to generate the simulation performance
results plotted in Figures 3 and 4. Figure 3 plots curves for the
number of time steps until failure versus the trial number. These
are typical curves for individual runs of the ASE/ACE learning con-
troller system. Figure 4 gives plots of the average number of trials
which are required to stabilize the cart plus pendulum system for
5000 time steps (100 seconds with $\Delta t = 0.02$ seconds). The average
number of trials are given as a function of the standard deviation
for the Gaussian normal random noise process with the mean for the
process as parameter. The mean equal zero curve indicates a trend
toward an increasing number of trials as the standard deviation is
increased from 0.01 to 0.25. The other curve for the mean equal to
0.10 shows relative constancy over the same range in standard devia-
tion. These runs were originally made to examine the sensitivity of
the ASE/ACE learning controller performance to variation in the
noise process used in generating its output function. An additional
objective was to develop a base from which the generalization and
robustness properties of the controller weights could be investi-
gated. Five runs were used to generate each point plotted in Figure
4. The results shown in Figures 3 and 4 are in general agreement
with those published by Barto et al [2]. However, the author has
found that his NRLNET131 implementation of the ASE/ACE controller
usually takes a lesser number of trials for successfully learning to
stabilize the cart plus pendulum system for both the 5000 Δt (100
seconds) cases shown here and the 200,000 Δt (66.7 minutes) cases
which the author ran to directly compare his results with those of
Barto et al. Extensive runs were not made for the 200,000 Δt (66.7
minutes) stabilization period because of the excessively long
elapsed time rerequired for the VAX system available to the author
to return answers for a single run.

Figure 5 shows the controller output force which is applied to
the cart in stabilizing the inverted pendulum. Here the applied
force is plotted as a function as a function of time over the first
100 Δt intervals (2 seconds). Extensive runs have been made with the
ASE/ACE controller system and all exhibit the characteristic +/- 10N
on-off or bang-bang behavior with $\Delta t = 0.02$ sec. This value of the
time increment should be adequate, based on physical system oscilla-
tion behavior, for the second order numerical integration scheme
used.

Figures 6-9 plot the four state variables: cart displacement
(x), cart velocity ($\dot{x}_c$), pendulum angular displacement (θ_p), and
pendulum angular velocity ($\dot{\theta}_p$), respectively. They are also plotted
as functions of time over the first 100 Δt intervals (2 seconds).
Consideration of these and similar time domain results for the state
variables and the applied force indicates that (a) significant in-
efficiencies can occur with respect to the input force and its im-
pact on the actual state variable behavior of the cart plus pendulum
system, (b) with $\Delta t = 0.02$ sec there may be some interaction be-
tween the numerical integration method used and the dynamics of the
ASE/ACE learning controller. To investigate (b) above, additional
runs were made in which Δt was reduced ($\Delta t = 0.01$, 0.005, 0.001
sec). These results although not included here did show significant
reduction with decreasing Δt in the bang-bang nature of the input

force and the higher frequency oscillations present in the state behavior over time (especially for the cart linear and the pendulum angular velocities).

The author has extended the single ASE/single ACE learning controller system to include two search and two critic elements. The elements in each pair work in parallel. Since the outputs are averaged in the 2 ASE/2 ACE learning controller system, it has -Fapp, 0, +Fapp as three possible outputs. This extension was implemented in the author's FORTRAN computer program NRLNET20. Initial runs indicate that the new controller as implemented has good performance up to a maximum learning point (maximum time for stability as a function of number of trials). Beyond this point the learning is severely degraded with increasing trials, or a form of limit cycle behavior occurs. These results indicate that the split-decision nature of the 2 ASE/2 ACE system as implemented in its averaging form may cause the observed behavior. In this case using a 3, 5, etc.(i.e., odd number of elements) in the ASE/ACE system may be warranted. These controllers would also have a "smoother" (i.e., less bang-bang) control action. Another alternative to improve performance is to more richly connect the elements both within and across the search element and the critic element layers. This would give the ASE/ACE neuronlike controller system a counter propagation/Grossberg layer plus Kohonen layer type of neural network structure [12, 13].

CONCLUSIONS

An examination has been made of the use of neural networks for the intelligent control of robotic arm plus hand/manipulator systems for the EVA Retriever. Based largely to the present time on a review of the literature and computer simulation work, this examination has indicated that a hierarchical, multi-layer neural network system can be used for intelligent control. Baseline feedforward control is used in conjunction with trajectory planning in these systems. Joint torque feedback provides the correction signal. These systems have the characteristic that as additional response behaviors are learned, much of the control action passes to the feedforward path.

Additional investigation into neural networks for intelligent control has focused on the use and extension of the Barto et al neuronlike ASE/ACE intelligent controller. A FORTRAN family of computer programs (NRLNETXX) were developed by the author as extensions of a previous Pascal language implementation of the controller at NASA/JSC. Work with these programs has concentrated on the following: (a) Verifying published results for convergence to stable solution (number of trials for a specified period of stability), (b) Developing graphics, etc. feedback tools to monitor system behavior (as, e.g., given by the applied control force and the four state variables as functions of time), (c) Investigate learning control behavior as a function of the number of unsupervised trials required to obtain stability and the random process parameters (Gaussian process mean and standard deviation), and (d) Basic extensions to the learning controller network incorporating two adaptive search elements (ASEs) and 2 adaptive critic elements (ACEs) in its structure.

RECOMMENDATIONS

This section of the report presents recommendations for the intelligent control of smart robotic arm plus hand systems using neural networks. These recommendations are based on the results presented above and on additional related work done by the author during Summer 1988. They are presented in the form of a research and development program plan. The R & D program plan gives activities that can continue the author's research begun during the 1988 summer program.

(1) Investigation of two dimensional graphics as a kinematic simulation tool for planning EVA object retrieval in terms of the approach to and grasping of objects using an articulated two-link arm/scissor hand system.

(2) Implementation of dynamics, sensing, and control models of the articulated two-link arm/scissor hand system. It is desired to mount this arm/hand system on a cart to represent the EVA Retriever in two dimensions.

(3) Examination of hierarchical neural networks with fuzzy logic reasoning as adaptive/general learning systems comprised of (a) Network architectures, (b) Transfer functions, and (c) Dynamic learning rules. These systems can employ joint torque and state vector feedback to control the arm/hand system(s) in object retrieval as discussed above.

(4) Investigation of extensions to the Barto/Anderson neuronlike learning system and counter propagation/back propagation type networks to the related problem of stabilizing/controlling the motion of simple and compound (articulated linkage) pendulums on a cart. Successful employment here can lead to similar use with the arm/hand retriever systems.

REFERENCES

1. "Extravehicular Activity (EVA) Retriever Program Plan," NASA/ Johnson Space Center, Houston, Texas, Report No. JSC-22144, May 1987.

2. A.G. Barto, R.S. Sutton, C.W. Anderson, "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13, 1983, pp 834-846.

3. A.G. Barto, C.W. Anderson, R.S. Sutton, "Synthesis of Non-linear Control Surfaces by a Layered Associative Search Network," Biological Cybernetics, Vol. 43, 1982, pp 175-185.

4. A.G. Barto, R.S. Sutton, "Landmark Learning: An Illustration of Associative Search," Biological Cybernetics, Vol. 42, 1981, pp 1-8.

5. A.G. Barto, R.S. Sutton, "Simulation of Anticipatory Responses in Classical Conditioning by a Neuronlike Adaptive Element," Behavioral Brain Research, Vol. 4, 1982, pp 221-235.

6. A.G. Barto, R.S. Sutton, P.S. Brouwer, "Associative Search

Network: A Reinforcement Learning Associative Memory,"
Biological Cybernetics, Vol. 40, 1981, pp 201-211.

7. Behnam Bavarian, "Introduction to Neural Networks for Intelligent Control," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 3-7.

8. Robert H. Cannon, Jr., Dynamics of Physical Systems, McGraw-Hill, New York, 1967, pp 703-710.

9. M. Caudill, "Neural Networks Primer: Parts I, II, and III," AI Expert, Part I, Dec. 1987, pp 46-52, Part II, Feb. 1988, pp 55-61, Part III, June 1988, pp 53-59.

10. M. Fujita, "Adaptive Filter Model of the Cerebellum," Biological Cybernetics, Vol. 45, 1982, pp 195-206.

11. Allon Guez, James L. Eilbert, Moshe Kam, "Neural Network Architecture for Control," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 22-25.

12. Robert Hecht-Nielsen. "Counterpropagation Networks," Proceedings, IEEE International Conference on Neural Networks, 1987.

13. Robert Hecht-Nielsen, "Applications of Counterpropagation Networks," Neural Networks, Volume 1, 1988.

14. J.J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proceedings of the National Academy of Sciences, Vol. 79, April 1982, pp 2554-2558.

15. Mitsuo Kawato, Yoji Uno, Michiaki Isobe, Ryoji Suzuki, "Hierarchical Neural Network Model for Voluntary Movement With Application to Robotics," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 8-16.

16. Mitsuo Kawato, "Adaptation and Learning in Control of Voluntary Movement in the Central Nervous System," Advanced Robotics, Vol. 2 (In Press).

17. M. Kawato, M. Etoh, Y. Oda, N. Tsukahara, "A New Algorithm for Voltage Clamp Iteration: A Learning Control of a Nonlinear Neuronal System," Biological Cybernetics, Vol. 53, 1985, pp 57-66.

18. M. Kawato, M. Isobe, Y. Maeda, R. Suzuki, "Coordinate Transformation and Learning control for Visually-Guided Voluntary Movement With Iteration: A Newton-Like Method in Function Space," Biological Cybernetics (In Press).

19. A.H. Klopf, "Brain Function and Adaptive Systems - A Heterostatic Theory," Air force Cambridge Research Laboratory Research Report AFCRL-72-0164, Bedford, Massachusetts, 1972. (A Summary Appears in Proceedings of the International Conference on Systems, Man, and Cybernetics, 1974).

20. A.H. Klopf, The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence, Hemisphere, Washington, D.C., 1982.

21. Daniel S. Levine, "Neural Networks: Growth Field for the 1990s," SIAM News, Vol. 21, No. 4, July 1988, pp 8-9.

22. Richard P. Lippmann, "An Introduction to Computing With Neural Nets," IEEE ASSP Magazine, IEEE Acoustics, Speech, and Signal Processing Society, Volume 4, Number 2, April 1987, pp 4-22.

23. R.A. McLauchlan, "Dynamics, Control and Sensor Issues Pertinent to Robotic Hands for the EVA Retriever System," Participant's

Network: A Reinforcement Learning Associative Memory,"
Biological Cybernetics, Vol. 40, 1981, pp 201-211.

7. Behnam Bavarian, "Introduction to Neural Networks for Intelligent Control," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 3-7.

8. Robert H. Cannon, Jr., Dynamics of Physical Systems, McGraw-Hill, New York, 1967, pp 703-710.

9. M. Caudill, "Neural Networks Primer: Parts I, II, and III," AI Expert, Part I, Dec. 1987, pp 46-52, Part II, Feb. 1988, pp 55-61, Part III, June 1988, pp 53-59.

10. M. Fujita, "Adaptive Filter Model of the Cerebellum," Biological Cybernetics, Vol. 45, 1982, pp 195-206.

11. Allon Guez, James L. Eilbert, Moshe Kam, "Neural Network Architecture for Control," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 22-25.

12. Robert Hecht-Nielsen. "Counterpropagation Networks," Proceedings, IEEE International Conference on Neural Networks, 1987.

13. Robert Hecht-Nielsen, "Applications of Counterpropagation Networks," Neural Networks, Volume 1, 1988.

14. J.J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proceedings of the National Academy of Sciences, Vol. 79, April 1982, pp 2554-2558.

15. Mitsuo Kawato, Yoji Uno, Michiaki Isobe, Ryoji Suzuki, "Hierarchical Neural Network Model for Voluntary Movement With Application to Robotics," IEEE Control Systems Magazine, Vol. 8, No. 2, April 1988, pp 8-16.

16. Mitsuo Kawato, "Adaptation and Learning in Control of Voluntary Movement in the Central Nervous System," Advanced Robotics, Vol. 2 (In Press).

17. M. Kawato, M. Etoh, Y. Oda, N. Tsukahara, "A New Algorithm for Voltage Clamp Iteration: A Learning Control of a Nonlinear Neuronal System," Biological Cybernetics, Vol. 53, 1985, pp 57-66.

18. M. Kawato, M. Isobe, Y. Maeda, R. Suzuki, "Coordinate Transformation and Learning control for Visually-Guided Voluntary Movement With Iteration: A Newton-Like Method in Function Space," Biological Cybernetics (In Press).

19. A.H. Klopf, "Brain Function and Adaptive Systems - A Heterostatic Theory," Air force Cambridge Research Laboratory Research Report AFCRL-72-0164, Bedford, Massachusetts, 1972. (A Summary Appears in Proceedings of the International Conference on Systems, Man, and Cybernetics, 1974).

20. A.H. Klopf, The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence, Hemisphere, Washington, D.C., 1982.

21. Daniel S. Levine, "Neural Networks: Growth Field for the 1990s," SIAM News, Vol. 21, No. 4, July 1988, pp 8-9.

22. Richard P. Lippmann, "An Introduction to Computing With Neural Nets," IEEE ASSP Magazine, IEEE Acoustics, Speech, and Signal Processing Society, Volume 4, Number 2, April 1987, pp 4-22.

23. R.A. McLauchlan, "Dynamics, Control and Sensor Issues Pertinent to Robotic Hands for the EVA Retriever System," Participant's

Final Report - 1987 NASA/ASEE Summer Faculty Research Program,
Department of Civil and Mechanical Engineering, Texas A & I
University, 3 September 1987.

24. D. Michie, R.A. Chambers, "BOXES: An Experiment in Adaptive
Control," in Machine Intelligence 2, E. Dale and D. Michie,
Editors. Oliver and Boyd, Edinburgh, 1968, pp 137-152.

25. D. Michie, R.A. Chambers, "BOXES as a Model of Pattern-Forma-
tion," in Toward a Theoretical Biology, Vol. 1, Prolegomena,
C.H. Waddington, Editor. Edinburgh University Press, Edin-
burgh, 1968, pp 206-212.

26. W.T. Miller, et al, "Application of a General Learning Algorithm
to the Control of Robotic Manipulators," International Journal
of Robotics Research, Vol. 6, No. 2, Summer 1987, pp 84-98.

27. Demetri Psaltis, Athanasios Sideris, Alan A. Yamamura, "Neural
Controllers," Proceedings IEEE First Conference on Neural Net-
works, San Diego, California, Vol. 4, June 1987, pp 551-558.

28. Demetri Psaltis, Athanasios Sideris, Alan A. Yamamura, "A
Multilayered Neural Network Controller," IEEE Control Systems
Magazine, Vol. 8, No. 2, April 1988, pp 17-21.

29. D.E. Rumelhart, J.L. McClelland, Parallel Distributed Proces-
sing: Explorations in the Microstructure of Cognition, Volumes
I and II, MIT Press, Cambridge, Massachusetts, 1986.

30. R.S. Sutton, A.G. Barto, "Toward a Modern Theory of Adaptive
Networks: Expectation and Prediction," Psychology Review, Vol.
88, 1981, pp 135-171.

31. R.S. Sutton, A.G. Barto, "An Adaptive Network That Constructs
and Uses an Internal Model of Its World," Cognition and Brain
Theory, Vol. 4, 1981, pp 213-246.

32. David W. Tank, John J. Hopfield, "Collective Computation in
Neuronlike Circuits," Scientific American, Volume 256, Number
12, December 1987, pp 104-114.

33. B. Widrow, M.E. Hoff, "Adaptive Switching Circuits," in 1960
WESCON Convention Record, Part IV, 1960, pp 96-104.

34. B. Widrow, N.K. Gupta, S. Maitra, "Punish/Reward: Learning
With a Critic in Adaptive Threshold Systems," IEEE Transac-
tions on Systems, Man, and Cybernetics, Vol. SMC-3, 1972,
pp 455-465.

35. B. Widrow, F.W. Smith, "Pattern-Recognizing Control Systems,"
in Computer and Information Sciences, J.T. Tow, R.H. Wilcox,
Editors. Clever Hume Press, 1964, pp 288-317.

36. B. Widrow, J.M. McCool, M.G. Larimore, and C.R. Johnson,
"Stationary and Nonstationary Learning Characteristics of the
LMS Adaptive Filter," Proceedings of the IEEE, Vol. 64, 1976,
pp 1151-1162.

MALADJUSTMENT OF KIDNEYS TO MICROGRAVITY:
DESIGN OF MEASURES TO REDUCE THE LOSS OF CALCIUM

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared By: Bohdan R. Nechay, D.V.M.

Academic Rank: Professor

University & Department: University of Texas Medical Branch
                                      Department of Pharmacology & Toxicology
                                      Galveston, Texas 77550

NASA/JSC
Directorate: Space and Life Sciences

Division: Medical Sciences

Branch: Biomedical Laboratories

JSC Colleague: Donald M. Foster, Ph.D.

Date Submitted: October 25, 1988

Contract Number: NGT 44-005-803

ABSTRACT

Losses of skeletal calcium and body fluids occur during prolonged exposure to microgravity. The kidney plays a major role in regulating the physiological functions involved. Relative to this regulatory function, the kidney performs three operations: filtration of blood plasma through the glomeruli, reabsorption, and secretion of fluid and electrolytes so that needed components are retained and only waste is eliminated in the urine.

Using data published in "Biomedical Results from Skylab" (1), we performed new calculations that reflect more directly the operations of the kidney in the handling of calcium, sodium, chloride, potassium and phosphate during space flight. These calculations revealed that the fraction of filtered calcium that was rejected by renal tubules and excreted in the urine increased by 71%, from 1.77% (preflight) to 3.02% (inflight) of the filtered load. This represents a large absolute increase because the total filtered amount is huge. Because the tubular rejection fraction of other ions increased relatively less than that of calcium, we postulate the inflight development of a specific renal defect that causes an excessive loss of calcium in urine and thereby contributes to the weakening of bones.

Because the reabsorption of glomerular filtrate (the conservation of body fluids and electrolytes by the kidney) depends upon the energy of adenosinetriphosphate (ATP) and enzymes that liberate its energy, we propose that an induction of the major renal enzyme system involved, sodium and potassium-dependent adenosinetriphosphatase, may improve the reabsorption of sodium directly and calcium indirectly and thereby lead to improved hydration and reduced calcium loss.

INTRODUCTION

Because of the known detrimental shifts of fluids and electrolytes, particularly calcium losses, that occur under conditions of microgravity, we undertook an evaluation of renal function studies during the Skylab flights and of renal handling of important ions. Using data published in "Biomedical Results from Skylab" (1), we performed new calculations that reflect more directly the operations of the kidneys of astronauts under conditions of microgravity.

To make this report clear to the nonspecialist, we now give a brief summary of pertinent renal physiology, in part derived from Pitts (2). The two kidneys weigh about 300 grams and thus constitute 0.4% of total body weight. In humans, each kidney is composed of about one million units called nephrons, which function in parallel. Each nephron begins with a glomerulus (microscopic sieve) attached to a tubule. Fluid filtered by the glomeruli passes through the renal tubules, to collecting ducts and eventually to the urinary bladder. The kidneys receive the greatest blood flow, in proportion to weight, of any organ of the body. Renal blood flow is equal to 20-25% of cardiac output or about 2000 liters per day. Urine formation starts with the filtration of blood plasma through the glomeruli at the rate of about 200 liters per day. The renal tubules reabsorb needed components of the glomerular filtrate, which contain most constituents of blood plasma excluding protein, and add to the filtrate undesirable products by secretion. The net result is excretion of waste in urine at an average rate of two liters per day. Thus, about 99% of glomerular filtrate is reabsorbed. Given the fact that an average adult has 40 liters of fluid in the body, the internal autodialysis of all body fluids through the kidneys takes place five times every 24 hours (200 liters of glomerular filtrate/40=5). The net result of renal filtration, reabsorption and secretion is the regulation of composition, within narrow limits compatible with life, of body fluids, often referred to as the constant internal environment.

METHODS

Glomerular filtration rates (GFR) and renal handling of ions were estimated (2,3) based on plasma and urinary concentrations obtained from published data from the nine crewmen who participated in Skylab flights (Tables 1 and 2). Where appropriate, plasma concentrations of substances obtained at different times were averaged to match the periods of urinary excretion data available.

The renal clearance of a substance is defined as the number of milliliters of plasma completely cleared of that substance in one minute. If the substance is completely filterable through glomeruli, that is, it is neither reabsorbed nor secreted, if it has no effect on renal function, and if it can be accurately measured in plasma and urine, then its clearance can be taken as a valid measure of GFR. Normal GFR per 1.73 square meters of body surface area, standard for an adult, averages 125 ml/min in men and 110 ml/min in women. In healthy persons, GFR is remarkably stable from day to day over a period of years.

The formula for calculation of renal clearance is U x V/P, where U and P represent concentrations of the substance in urine and plasma (mg/ml) respectively, and V represents urine formation rate in ml/min. If the clearance of a given substance is larger than the simultaneous GFR, that substance is said to be "secreted" by the tubules of the kidney. If clearance of a substance is smaller than GFR, the substance is said to

be "reabsorbed." It should be noted that, in the absence of more detailed studies, the clearance of a substance is a net figure and does not take into account simultaneous secretion and reabsorption that might also be taking place.

The standard method of determination of GFR is by measurement of inulin clearance following intravenous infusion. Inulin is a high (5,000) molecular weight polysaccharide that is completely filterable and is neither secreted nor absorbed by the kidney. Obviously, this method cannot be used during space flight. Instead, measurement of endogenous creatinine clearance, which agrees reasonably well with the inulin clearance, is usually taken as an approximate measure of GFR and was used in our calculations. Creatinine is a product of muscle creatine phosphate; its production is constant and is proportional to muscle mass. Renal elimination of creatinine is also constant. The agreement of GFR as measured by inulin and by creatinine clearances results from the balance of two errors: 1) the analysis method slightly overestimates the creatinine concentration in plasma (but not in urine), and this gives the appearance of a lower creatinine clearance; and 2) the true clearance of creatinine is slightly higher than GFR because of tubular secretion. These are the limitations in using creatinine clearance as an approximate measure of GFR in these studies.

Filtration rate of an ion (mg/min) through glomeruli was calculated by multiplying its filterable concentration in plasma (mg/ml) by the GFR (ml/min). Because 40% of plasma calcium is bound to plasma proteins, which are too large to pass through the glomeruli, the filterable fraction of calcium was taken as 60% of total plasma calcium. Phosphate was assumed 90% filterable. Virtually all plasma sodium, potassium and chloride are freely filterable.

The tubular reabsorption rate (mg/min) of an ion is its filtration rate (mg/min) minus its excretion rate (mg/min). Tubular rejection fraction or fractional excretion (%) was calculated by the formula: Excretion rate (mg/min)/ filtration rate (mg/min) x 100. This yields the percentage of the filtered quantity of a given ion that escaped tubular reabsorption.

RESULTS

Renal function data calculated from Tables 1 and 2 are shown in Table 3. Creatinine clearances proved remarkably stable throughout the missions, indicating intact renal vascular autoregulatory mechanisms that control the GFR in the face of the well known circulatory disturbances and body fluid shifts that occur during space flight (4-6). In flight, GFR's held within 2-3% of the preflight values and fell only a modest 9% during the 60-85 day flight period.

Tubular reabsorption of individual ions was depressed during flight and reverted postflight to baseline, preflight values (Fig. 1). Peak fractional excretion rose 71% (from 1.77% to 3.02%) for calcium, 12% for phosphate, 38% for chloride and 40% for sodium. This indicates a generalized increased tubular rejection of several ionic components of the glomerular filtrate with a relative selectivity for calcium.

The data for renal handling of calcium are plotted in Figure 2. The curves for fractional excretion and urinary excretion rate of calcium were similar: the former exhibited a smoother natural progression, suggesting that this was the mechanism for increased calcium loss in the urine. Increased fractional excretion coincided with a

slight but significant rise in plasma calcium concentration (peak 6%). Filtered and reabsorbed calcium increased (peak 8-9%) during flight and paralleled the creatinine clearance curve throughout. Thus, the increased fractional excretion of calcium during flight was the dominant finding.

To see more clearly the mechanism of tubular transport of calcium, in Figure 3 we have plotted renal calcium handling adjusted per 100 ml of GFR. The curve for the urinary calcium excretion rate resembled more closely the curve for fractional calcium excretion than that shown in Figure 2. The adjusted filtration and reabsorption rates of calcium remained parallel but became more stable than those shown in Figure 2, and peak increases during flight were only 4-5%. This is further evidence that the calcium excreted in urine appears there as a result of diminished tubular reabsorption.

The urinary excretion of 0.29 gm of calcium per day during the Skylab flights was nearly twice that of control values of 0.16 gm per day. Thus, 7.8 to 11.05 gm of calcium were lost during 59 day and 84 day Skylab flights respectively. These amounts are equal to more than the total of extraskeletal intracellular and extracellular calcium stores.

DISCUSSION

The outstanding finding of our study is that renal tubular reabsorption of calcium abruptly diminished during space flight and rapidly returned to control levels after landing. This resulted in urinary excretion of 0.29 gm of calcium per day during flight, up from 0.16 gm per day control rate. The net total calcium loss attributable to the kidney during the Skylab flights of 59 to 84 days equaled 7.8 gm to 11.05 gm respectively. The crewmen had an adequate average dietary intake of 0.73 gm calcium per day but went into negative calcium balance because fecal calcium excretion also increased (1,7). The calcium-wasting by the kidney is further demonstrated by the lack of renal conservation of calcium in the face of the increased fecal calcium losses.

To place the magnitude of the calcium loss in perspective, it is essential to realize that all extracellular body fluids contain only about 1 gm of calcium, cells contain 5 gm, and bone contains a total of 1,200 gm (3). Thus, any sizeable calcium loss, like that which occurred during the Skylab flights, must come eventually from bone. Calcium homeostasis involves continuous mineral turnover in bone, absorption from food and reabsorption by the kidney. Overall calcium metabolism is primarily set to maintain a constant calcium concentration in extracellular fluid, even at the expense of bone density. Absorption from the gut is controlled by vitamin D, bone resorption and renal reabsorption are promoted by parathyroid hormone and inhibited by calcitonin (3,8).

Based on the above considerations, options for prevention of calcium loss are: 1) increase intake and/or decrease losses of calcium from the gut; 2) decrease calcium loss from bone; and 3) decrease calcium loss from the kidney. Relative to option 1, stable and adequate calcium intake during Skylab flights did not prevent negative calcium balance (1,7). In fact, there was an increased amount of calcium lost in feces (7). Enhancement of calcium absorption from food deserves further study.

Decreased egress of calcium from bone appears to be under most intensive investigation at present as a solution to bone demineralization during space flight. In our opinion, producing a halt to bone resorption without increased absorption from the gut and reduced loss via the kidneys would lead to a reduced intracellular and

extracellular calcium concentration. Since calcium is required for muscular contraction and nerve conduction, inadequate calcium levels may lead to increased weakness and atrophy, already a problem during space flight (9). Diminished supply of calcium to the heart could lead to impairment of cardiac output and to irregularities of heart rhythm (arrhythmias) that could result in sudden death.

Relative to the third option, reversal of the acquired defect in tubular reabsorption of calcium would be a desirable solution to the excessive calcium loss during space flight. Because renal calcium reabsorption is in part dependent upon the transport of sodium, we propose that an induction of renal sodium and potassium-dependent adenosinetriphosphatase (Na and K ATPase), an enzyme system that allows the use of adenosine triphosphate (ATP) for ion transport (10), would improve tubular reabsorption of sodium directly and that of calcium secondarily. In addition to calcium conservation, this maneuver would improve reabsorption of sodium, chloride and water, all of which are also excessively lost during space flight. Methods for induction of renal Na and K ATPase deserve further investigation in an attempt to prevent osteoporosis and fluid and electrolyte loss associated with space travel. The magnitude of renal-related calcium losses, coupled with the fact that Skylab astronauts had adequate calcium intake and excessive fecal calcium losses, make it wise to consider ways in which renal losses of calcium can be curtailed.

REFERENCES

1. Leach, C.S. & Rambaut, P.C. Biomedical responses of the skylab crewmen: an overview. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein (Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 204-216, 1977.

2. Pitts, R.F. Physiology of the kidney and body fluids, Year Book Medical Publishers, Chicago, Ill., 1974.

3. Sutton, R.A.L. & Dirks, J.H. Calcium and magnesium: Renal handling and disorders of metabolism. In The Kidney, Vol. 1, B.M. Brenner and F.C. Rector, Jr. (Eds.), W. B. Saunders Co., Philadelphia, Pa., pp. 551-618, 1986.

4. Henry, W.L., Epstein, S.E., Griffith, J.M., Goldstein, R.E., & Redwood, D.R. Effect of prolonged space flight on cardiac function and dimensions. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein (Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 366-371, 1977.

5. Johnson, R.L., Hoffler, G.W., Nicogossian, A.E., Bergman, S.A., & Jackson, M.M. Lower body negative pressure: Third manned Skylab mission. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein (Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 284-312, 1977.

6. Thornton, W.E., Hoffler, G.W., & Rummel, J.A. Anthropometric changes and fluid shifts. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein (Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 330-338, 1977.

7. Whedon, G.D., Lutwak, L., Rambaut, P.C., Whittle, M.W., Smith, M.C., Reid, J., Leach, C., Stadler, D.R., & Sanford, D.D. Mineral and nitrogen metabolic studies, experiment M071. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein

(Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 164-174, 1977.

8. Coburn, J.W. & Slatopolsky, E. Vitamin D, parathyroid hormone and renal osteodystrophy. In The Kidney, Vol. 2, B.M. Brenner and F.C. Rector, Jr. (Eds.), W. B. Saunders Co., Philadelphia, Pa., pp. 1657-1729, 1986.

9. Thornton, W.E. & Rummel, J.A. Muscular deconditioning and its prevention in space flight. In Biomedical Results from Skylab, R.S. Johnston and L.F. Dietlein (Eds.), National Aeronautics and Space Administration, Washington, D.C., pp. 191-197, 1977.

10. Nechay, B.R. Biochemical basis of diuretic action. J. Clin. Pharmacol. 17:626-641, 1977.

TABLE 1.-FROM REFERENCE 1.

TABLE 23-IX.—*Skylab Summary, Urine Biochemical Results (9 Crewmen)*

(Mean ± Standard error)

Units	Measured substance	Preflight day	In-flight day			Postflight day		
			1-28	29-59	60-85	1-6	7-13	14-18
meq/TV	Sodium	160.0 ± 3.0	[1]174.0 ± 3.0	[1]190.0 ± 7.0	[1]199.0 ± 6.0	[1]121.0 ± 11.0	[1]170.0 ± 6.0	[1]173.0 ± 11.0
meq/TV	Potassium	74.0 ± 1.0	[1]82.0 ± 2.0	[1]80.0 ± 2.0	[1]81.0 ± 3.0	[1]65.0 ± 4.0	76.0 ± 4.0	82.0 ± 5.0
meq/TV	Chloride	148.0 ± 4.0	[1]162.0 ± 5.0	[1]177.0 ± 6.0	[1]180.0 ± 5.0	[1]116.0 ± 11.0	[1]160.0 ± 6.0	164.0 ± 11.0
mg/TV	Creatinine	1955.0 ± 20.0	2079.0 ± 40.0	2104.0 ± 55.0	2081.0 ± 31.0	2005.0 ± 95.0	2037.0 ± 78.0	1969.0 ± 109.0
mOsmoles	Osmolality	650.0 ± 17.0	[1]789.0 ± 27.0	[1]791.0 ± 19.0	[1]717.0 ± 24.0	593.0 ± 60.0	[1]549.0 ± 49.0	584.0 ± 66.0
meq/TV	Calcium	8.0 ± 0.2	[1]14.4 ± 0.8	[1]14.5 ± 0.8	[1]11.8 ± 0.4	[1]11.2 ± 1.6	8.8 ± 1.0	8.3 ± 1.0
mg/TV	Phosphates	1045.0 ± 15.0	[1]1270.0 ± 27.0	[1]1196.0 ± 35.0	[1]1181.0 ± 30.0	[1]934.0 ± 55.0	1029.0 ± 55.0	1031.0 ± 50.0
mg/TV	Uric Acid	969.0 ± 15.0	[1]899.0 ± 22.0	[1]934.0 ± 38.0	[1]884.0 ± 33.0	[1]884.0 ± 41.0	929.0 ± 50.0	942.0 ± 53.0
meq/TV	Magnesium	8.9 ± 0.1	[1]10.8 ± 0.2	[1]9.4 ± 0.4	8.7 ± 0.5	[1]7.7 ± 0.5	9.1 ± 0.4	9.1 ± 0.4
μg/TV	Cortisol	54.3 ± 4.1	[1]94.4 ± 4.8	[1]83.6 ± 4.0	[1]90.2 ± 5.3	[1]69.5 ± 5.8	[1]63.3 ± 6.0	[1]76.6 ± 8.0
μg/TV	Aldosterone	11.3 ± 1.1	[1]32.8 ± 2.2	[1]22.4 ± 1.7	[1]30.0 ± 3.1	[1]18.6 ± 4.3	11.8 ± 3.0	11.4 ± 3.3
μg/TV	Epinephrine	27.2 ± 4.6	24.8 ± 1.4	21.3 ± 1.7	38.1 ± 3.3	37.2 ± 3.1	33.7 ± 3.4	37.5 ± 7.2
μg/TV	Norepinephrine	69.4 ± 6.0	59.9 ± 2.0	66.7 ± 4.0	65.2 ± 6.4	[1]99.4 ± 6.2	[1]88.8 ± 6.4	[1]89.6 ± 6.6
mμ/TV	Antidiuretic hormone	50.3 ± 10.0	41.9 ± 4.3	[1]24.1 ± 2.4	[1]20.3 ± 2.5	46.5 ± 10.0	[1]25.6 ± 8.0	31.0 ± 8.2
mg/TV	Total 17 Hydroxy-corticosteroids	6.1 ± 0.4	6.2 ± 0.4	6.5 ± 0.3	6.2 ± 1.0	[1]5.2 ± 0.5	[1]5.1 ± 0.4	[1]5.2 ± 0.8
mg/TV	Total 17 Ketosteroids	7.0 ± 0.5	[1]10.3 ± 0.4	[1]10.8 ± 0.5	[1]13.5 ± 1.3	7.0 ± 0.7	7.4 ± 0.5	7.6 ± 0.6

[1] $P \leq 0.05$.

TV Total volume 24h

TABLE 2.—FROM REFERENCE 1.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 23–V.—*Skylab Summary, Plasma Biochemical Results (9 Crewmen)*

(Mean ± Standard error)

No.		Sodium[1]	Potassium	Chloride	Creatinine	Glucose	Osmolality	Calcium	Phosphate
		meq/liter	meq/liter	meq/liter	mg pct.	mg pct.	mOsmoles	mg pct.	mg pct.
36	Preflight	141 ± 0.7	4.12 ± 0.04	97.7 ± 0.5	1.26 ± 0.03	86.6 ± 0.03	290 ± 0.8	9.7 ± 0.05	3.4 ± 0.1
	Mission day								
9	3, 4	139 ± 2	4.26 ± 0.08	96.8 ± 0.7	1.31 ± 0.03	90.3 ± 2.4	289 ± 1	*10.4 ± 0.1	3.7 ± 0.3
8	5, 6	*137 ± 2	4.30 ± 0.14	96.9 ± 0.8	1.27 ± 0.03	86.7 ± 1.8	*287 ± 1	*10.2 ± 0.1	*3.6 ± 0.3
6	13, 14	137 ± 1	4.41 ± 0.15	*94.7 ± 1.1	1.28 ± 0.03	86.7 ± 1.8	286 ± 2	*10.2 ± 0.1	*3.9 ± 0.3
6	20, 21	140 ± 1	4.25 ± 0.11	95.7 ± 0.8	1.35 ± 0.03	87.0 ± 1.8	289 ± 2	*10.1 ± 0.2	*3.4 ± 0.1
6	27, 30	*138 ± 0.8	4.25 ± 0.10	*95.2 ± 0.8	1.27 ± 0.03	84.3 ± 2.3	*287 ± 2	*10.4 ± 0.1	*3.9 ± 0.3
6	38	*136 ± 2	*4.05 ± 0.15	93.5 ± 1.2	1.31 ± 0.07	*80.1 ± 2.5	*280 ± 4	10.1 ± 0.2	*3.1 ± 0.5
6	45, 48	*137 ± 2	4.30 ± 0.13	94.5 ± 0.7	1.34 ± 0.03	*84.4 ± 1.4	287 ± 3	*10.1 ± 0.1	*3.8 ± 0.1
6	58, 59	*137 ± 2	4.19 ± 0.13	94.0 ± 1.5	1.38 ± 0.12	*81.8 ± 2.2	286 ± 4	*10.1 ± 0.2	*3.8 ± 0.2
3	73	139 ± 2	3.75 ± 0.20	94.6 ± 1.2	1.51 ± 0.05	80.9 ± 2.2	284 ± 2	10.1 ± 0.3	*3.9 ± 0.2
3	82	137 ± 0.6	4.19 ± 0.06	95.8 ± 0.2	1.54 ± 0.03	*81.0 ± 1.2	*285 ± 2	10.1 ± 0.1	3.6 ± 0.1
	Recovery (R)								
9	R + 0	139 ± 1	4.18 ± 0.05	*96.2 ± 1.0	1.28 ± 0.05	*100.5 ± 2.6	289 ± 1	*10.0 ± 0.1	*3.9 ± 0.2
9	R + 1	139 ± 1	4.10 ± 0.08	*96.4 ± 1.0	1.31 ± 0.06	92.3 ± 2.8	289 ± 1	*10.1 ± 0.1	*3.6 ± 0.03
9	R + 3, 4	139 ± 1	4.02 ± 0.13	96.9 ± 1.0	1.26 ± 0.06	*90.5 ± 1.4	*294 ± 2	9.8 ± 0.1	3.4 ± 0.2
6	R + 14	141 ± 0.8	4.05 ± 0.05	97.7 ± 1.6	1.33 ± 0.09	85.4 ± 0.7	289 ± 2	*9.4 ± 0.1	2.8 ± 0.2

[1] Corrected for Na-EDTA.   * $P \leq 0.05$.

TABLE 23–VI.—*Skylab Summary, Plasma Biochemical Results (9 Crewmen)*

(Mean ± Standard error)

No.		Cortisol	Angiotensin I	Aldosterone	ACTH	Insulin	HGH	PTH
		µg/100 ml	ng/ml per hour	pg/100 ml	pg/ml	µU/ml	ng/ml	ng/ml
39	Preflight	12.2 ± 0.7	0.77 ± 0.14	180 ± 25	35.7 ± 3.3	17 ± 0.6	1.3 ± 0.2	17 ± 1
	Mission Day							
9	3, 4	12.7 ± 1.6	1.09 ± 0.24	176 ± 58	[1]15.2 ± 4.9	15 ± 2	[1]2.1 ± 0.5	17 ± 2
8	5, 6	[1]14.8 ± 1.0	1.75 ± 0.42	163 ± 75	26.5 ± 9.2	18 ± 6	1.2 ± 0.3	16 ± 3
6	13, 14	13.4 ± 1.7	.91 ± 0.28	252 ± 65	33.0 ± 8	18 ± 3	1.5 ± 0.2	14 ± 1
6	20, 21	12.3 ± 1.5	.52 ± 0.12	163 ± 90	[1]11.9 ± 4	[1]8 ± 1	1.2 ± 0.3	20 ± 4
6	27, 30	13.6 ± 2.1	.45 ± 0.16	204 ± 88	32.0 ± 7	20 ± 3	3.2 ± 2.0	14 ± 2
6	38	13.7 ± 1.0	.72 ± 0.36	94 ± 17	17.7 ± 11.6	[1]10 ± 1	1.1 ± 0.3	15 ± 2
6	45, 48	14.3 ± 1.3	.37 ± 0.10	118 ± 7	[1]12.1 ± 5.8	[1]9 ± 2	1.5 ± 0.5	18 ± 4
6	58, 59	[1]13.5 ± 0.7	[1]1.11 ± 0.51	148 ± 31	32.3 ± 18.7	[1]9 ± 2	1.6 ± 0.4	18 ± 3
3	73	14.5 ± 3.4	.27 ± 0.08	117 ± 39		9 ± 4	0.6 ± 0.1	24 ± 2
3	82	[1]16.1 ± 0.6	.32 ± 0.04	142 ± 17		11 ± 4	0.7 ± 0.1	25 ± 2
	Recovery (R)							
9	R + 0	13.2 ± 2.1	.71 ± 0.23	215 ± 74	23.8 ± 6.3	20 ± 3	[1]2.9 ± 0.6	17 ± 2
9	R + 1	10.8 ± 1.0	[1]2.15 ± 0.55	[1]478 ± 77	[1]24.0 ± 7.5	20 ± 2	[1]2.8 ± 0.8	19 ± 3
9	R + 3, 4	13.7 ± 3.0	.86 ± 0.45	[1]357 ± 65	[1]23.3 ± 2.4	18 ± 2	[1]2.6 ± 0.8	19 ± 3
9	R + 13, 14	10.6 ± 0.7	[1].14 ± 0.05	153 ± 35	38.2 ± 13.9	17 ± 3	1.2 ± 0.2	18 ± 4

[1] $P \leq 0.05$.

TABLE 3.-GLOMERULAR FILTRATION RATE (CREATININE CLEARANCE) AND RENAL ELECTROLYTE TRANSPORT IN 9 SKYLAB CREWMAN. DATA RE-CALCULATED FROM TABLES 1 AND 2.

| SUBSTANCE	FUNCTION	UNITS	PREFLIGHT DAY	IN FLIGHT DAYS			POSTFLIGHT DAYS		
				1-28	29-59	60-85	1-6	7-13	14-18
Creatinine	Urinary excretion rate (UV)	mg/min	1.3576	1.4438	1.4611	1.4451	1.3924	1.4146	1.3674
	Plasma concentration (P)	mg/ml	0.0126	0.0130	0.0133	0.0148	0.0128		0.0133
	Clearance (UV/P)	ml/min	108	111	110	98	109		103
Calcium	Plasma concentration (total)	mg/ml	0.0970	0.1030	0.1010	0.1010	0.1000		0.0940
	Plasma concentration (60%)	mg/ml	0.0582	0.0618	0.0606	0.0606	0.0600		0.0564
	Filtration rate	mg/min	6.2856	6.8598	6.6660	5.9388	6.5400		5.8092
	Reabsorption rate	mg/min	6.1745	6.6598	6.4646	5.7749	6.3844		5.6939
	Fractional Reabsorption	%	98.23	97.08	96.92	97.24	97.62		98.02
	Excretion rate	mg/min	0.1111	0.2000	0.2014	0.1639	0.1556	0.1222	0.1153
	Fractional excretion	%	1.77	2.92	3.02	2.76	2.38		1.98
Phosphate	Plasma concentration (total)	mg/ml	0.0340	0.0370	0.0345	0.0377	0.0362		0.0280
	Plasma concentration (90%)	mg/ml	0.0306	0.0333	0.0311	0.0339	0.0326		0.0252
	Filtration rate	mg/min	3.3048	3.6963	3.4210	3.3222	3.5534		2.5956
	Reabsorption rate	mg/min	2.5791	2.8144	2.5904	2.5021	2.9048		1.8796
	Fractional reabsorption	%	78.0	76.1	75.7	75.3			72.4
	Excretion rate	mg/min	0.7257	0.8819	0.8306	0.8201	0.6486	0.7146	0.7160
	Fractional excretion	%	22.0	23.9	24.3	24.7	18.3		27.6
Chloride	Plasma concentration	mM	97.7	95.9	94.0	94.8	96.5		97.7
	Filtration rate	µEq/min	10,552	10,645	10,340	9,290	10,519		9,991
	Reabsorption rate	µEq/min	10,449	10,532	10,217	9,165	10,438		9,877
	Fractional reabsorption	%	99.03	98.94	98.81	98.65	99.23		98.86
	Excretion rate	µEq/min	103	113	123	125	81	111	114
	Fractional excretion	%	0.97	1.06	1.19	1.35	0.77		1.14
Sodium	Plasma concentration	mM	141	138	137	138	139		141
	Filtration rate	µEq/min	15,228	15,318	15,070	13,524	15,151		14,523
	Reabsorption rate	µEq/min	15,117	15,197	14,937	13,386	15,067		14,403
	Fractional reabsorption	%	99.27	99.21	99.12	98.98	99.45		99.17
	Excretion rate	µEq/min	111	121	133	138	89	118	120
	Fractional excretion	%	0.73	0.79	0.88	1.02	0.55		0.83
Potassium	Plasma concentration	mM	4.12	4.29	4.18	4.04	4.10		4.05
	Filtration rate	µEq/min	445	477	459	396	447		417
	Excretion rate	µEq/min	51.4	56.9	55.6	56.3	45.1	52.8	56.9
	Fractional excretion	%	11.55	11.93	12.11	14.22	10.09		13.65
Sodium/ Potassium	Excretion rate/ Excretion rate	Ratio	2.16	2.12	2.38	2.45	1.86	2.25	2.11
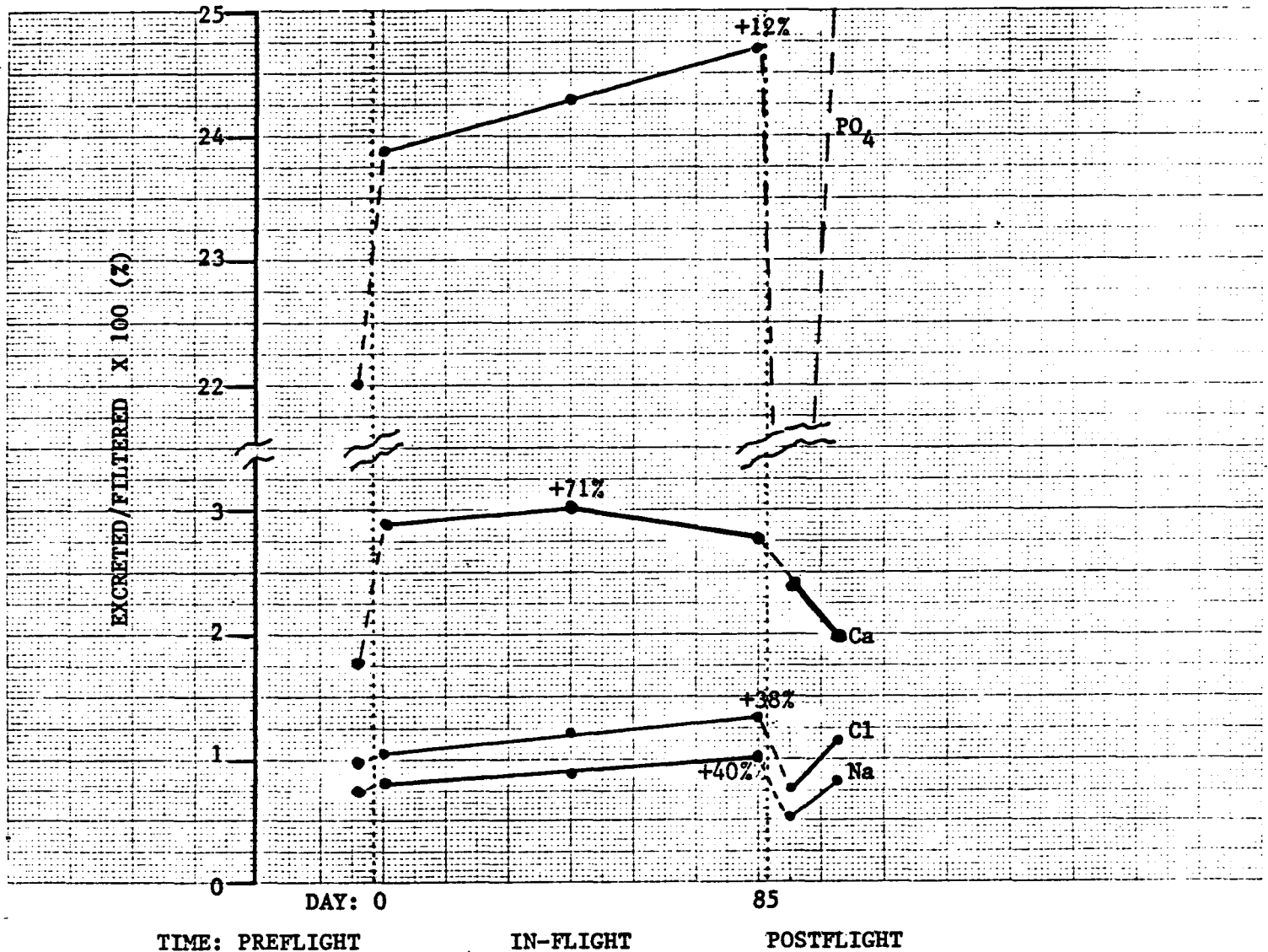
Figure 1. Renal function of nine Skylab crewmen: Fractional
excretion of calcium, phosphate, sodium, and chloride. The data
indicate filtered fractions of these ions that escaped reabsorp-
tion and were excreted in urine. Peak percent changes shown in
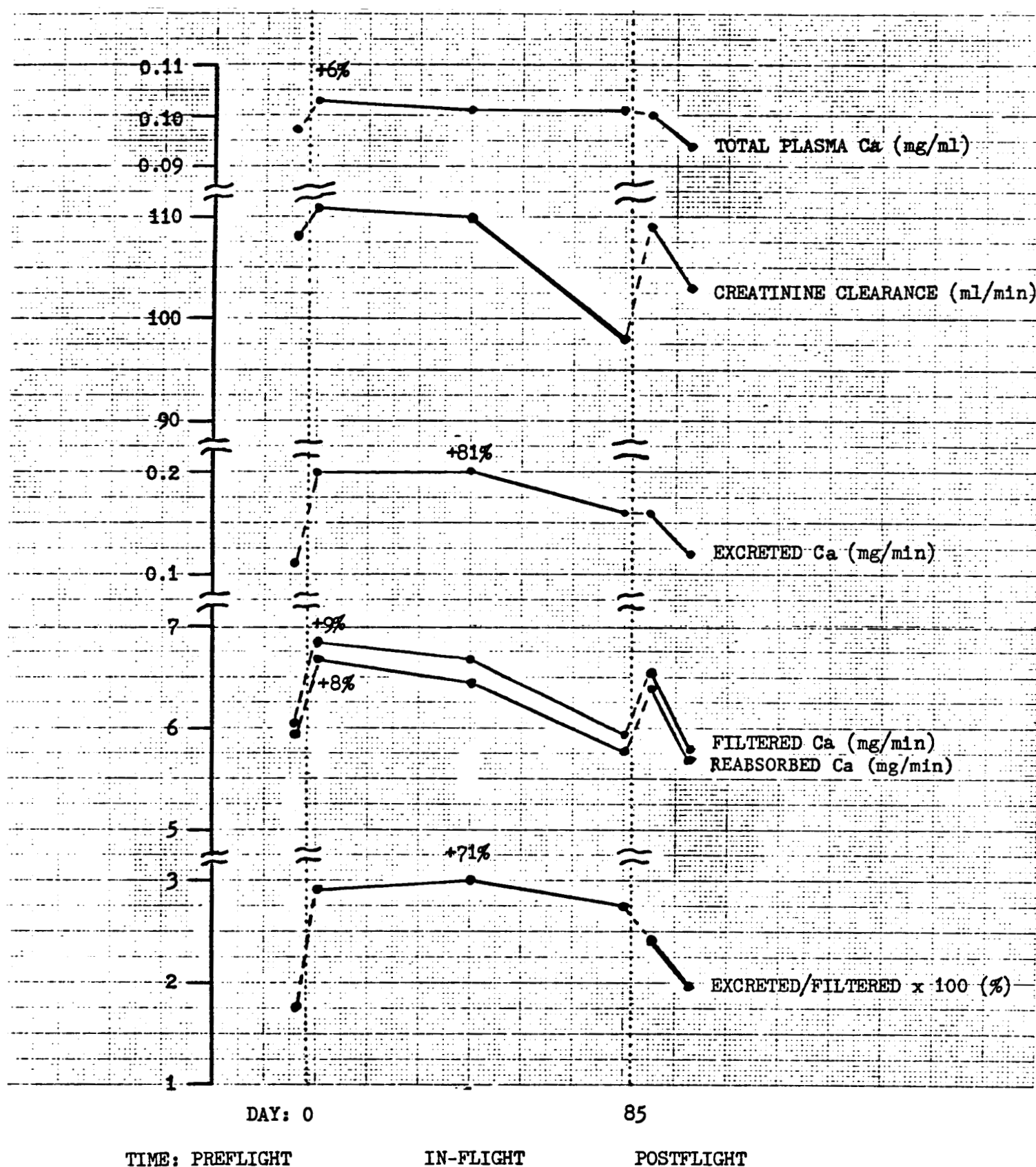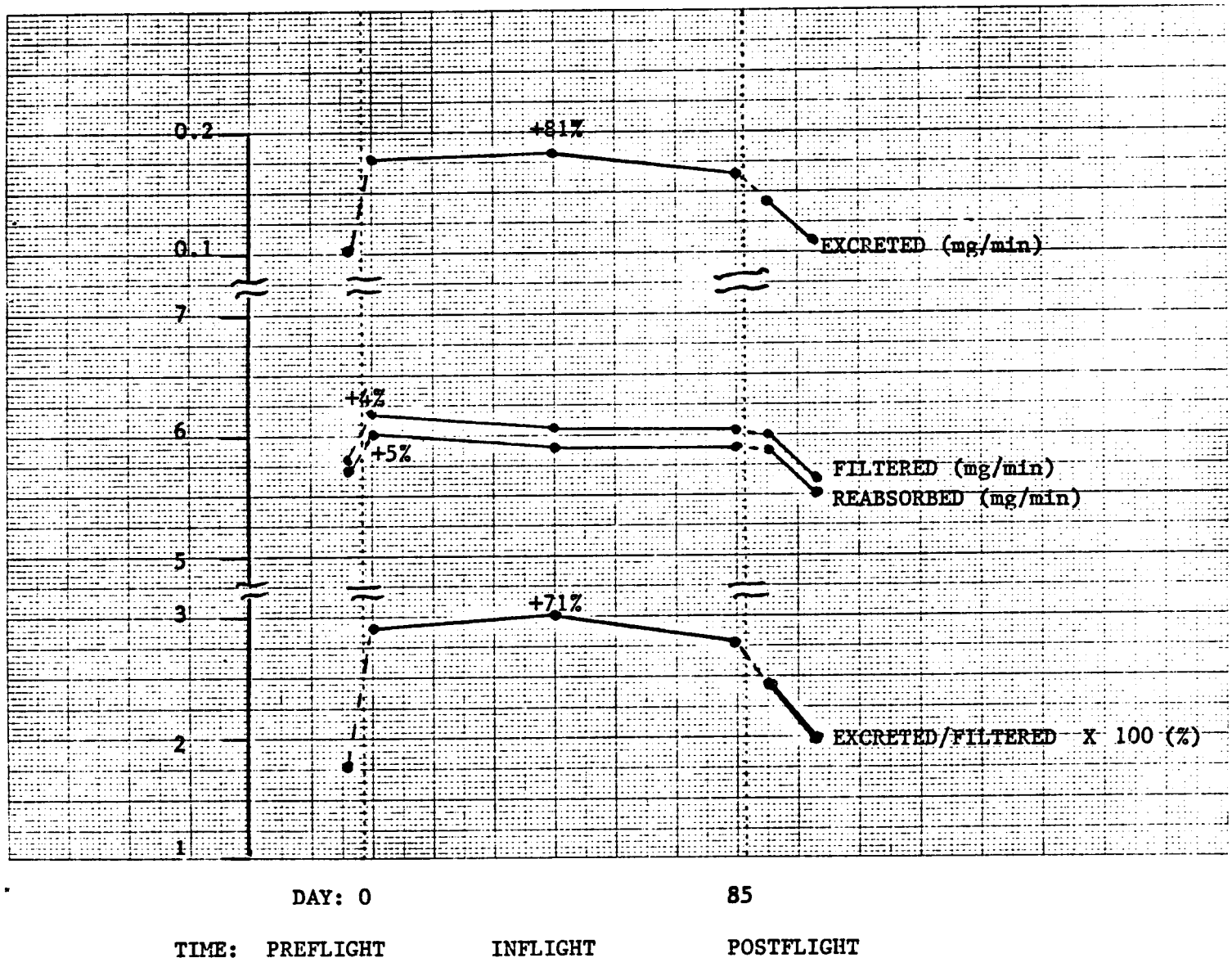the figure are relative to preflight values.

Figure 2. Plasma calcium concentration, GFR, and renal handling of calcium in nine Skylab crewmen. Peak percent changes indicated in the figure are relative to preflight values.

Figure 3. Renal handling of calcium per 100 ml/min GFR in nine
Skylab crewmen. Peak percent changes indicated in the figure are
relative to preflight values.

ACKNOWLEDGEMENT

To my wife, Karyl, a scientist and clinician, whose genuine interest and sharing of ideas made this project even more fun.

EVALUATION OF CRUSTAL RECYCLING DURING THE EVOLUTION
OF ARCHEAN-AGE MATACHEWAN BASALTIC MAGMAS

Final Report

NASA/JSC Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared By: Dennis O. Nelson, Ph.D.

Academic Rank: Professor

University & Department: Sul Ross State University
                         Geology Department
                         Alpine, Texas 79832

NASA/JSC

Directorate: Space and Life Sciences

Division: Solar System Exploration

Branch: Planetary Science

JSC Colleague: Donald A. Morrison, Ph.D.

Date Submitted: August 12, 1988

Contract Number: NGT 44-005-803

ABSTRACT

The Matachewan-Hearst Dike swarm consists of plagioclase megacryst-bearing tholeiite dikes that were emplaced over an area of approximately 250,000 km2 of the Superior Province near the end of the Archean (2.45 Ga). Invaded supracrustal rocks vary from metavolcanic dominated granitoid-greenstone belts to subprovinces dominated by metasedimentary lithologies. The dike compositions vary from depleted ([La/Sm]n < 0.8) to moderately enriched ([La/Sm]n > 1.80) and define linear arrays in compositional space (data can be obtained from the author). Neither the composition of the immediate country rock nor the individual subprovince in which a given dike is located exerts significant control on dike composition. Although variations in individual element concentrations were produced by shallow combined replenishment-fractional crystallization (RFC), the range in incompatible element ratios, e.g. [La/Sm]n, insensitive to RFC, requires additional processes. Relative abundances of the rare earth (REE) and high field strength (HFSE) elements suggest either crustal contamination or involvement of a slab-derived component. Distinguishing between these two alternatives is important in determining whether or not crustal formation/evolution at the end of the Archean was by subduction-related processes or some alternative such as underplating. Combined assimilation-fractional crystallization (AFC) models, using crustal rock data from xenoliths and from the adjacent Kapuskasing Structural Zone (KSZ) in which lower crustal rocks are exposed, suggest that assimilation can accommodate the incompatible element variation. The assimilant was dominated by silicic granitoids (tonalite to granodiorite) although mixing of parental dike magmas with partial melts of previously underplated basaltic rocks best explains some dikes. Assimilation rate, expressed as mass of assimilant to mass of cumulate phases, was approximately 0.5; total relative mass assimilated varied from 5 to 20 percent. Although the Matachewan-Hearst magmas possessed compositional signatures often attributed to a slab-derived component, e.g. high REE/HFSE, Ta/Ta* = 0.17-0.41 and falling within the 'destructive plate margin' field on a Th-Hf-Ta diagram, modeling indicates that these characteristics can be produced by AFC involving observed depleted dike compositions and granitoids of the KSZ. Partial melt models of a hypothetical amphibole-bearing peridotite (slab-component enriched [?]) produce variations in Ti/Ti* and (Tb/Yb)n that are not compatible with the MD data. AFC processes during underplating of the existing Archean crust produced the variations in incompatible element ratios, including the relative depletions of the HFSE.

18-2

INTRODUCTION

Basaltic magma, derived through the partial melting of planetary mantles, is a fundamental product of all the terrestrial planets (1). The ascent of basalt magma is one of the principal mechanisms of transfering `juvenile' mantle material to the surface. Subsequent chemical differentiation of basaltic magma in that environment is important in the generation of new crust and the modification of pre-existing crust. Data from missions to the other terrestrial planets (i.e. Mercury, Venus, Moon and Mars) suggests that all of these bodies have undergone differentiation into crust and mantle domains to varying extents; Earth and Moon are endmembers representing the greatest and least extents of this differentiation. Although there is broad agreement that the Earth's crust formed through chemical processing of the mantle and its derivatives, the mechanism(s) by which this was accomplished is(are) a matter of debate (2, 3). Although estimates vary, it is clear that the bulk of the Earth's crust formed prior to the Late Archean (2.5 billion years ago (Ga)) (3). Of interest is whether crustal growth and evolution during the Late Archean occurred in a manner similar to more modern environments, (e.g. involving lateral plate motions and in particular, subduction processes), or was early crustal genesis dominated by vertical tectonics, involving the process of underplating (e.g. 4). In the latter process, large volumes of basalt magma pool at the base of the low-density crust. The subsequent crystallization and differentiation of this magma adds to mass of the crust. Distinguishing between a modern plate tectonic model and the underplating process requires knowledge of the chemical nature of the mantle source (i.e. is it comparable to modern sources of convergent zone magmas) and the extent and nature of crustal recycling (i.e. crustal melting and assimilation) during the underplating process. In this particular study, the major and trace element compositions of dikes from the extensive Matachewan-Hearst swarm are evaluated in order to characterize their mantle source(s) and to identify the process(es) of magmatic differentiation that operated during their evolution.

BACKGROUND

The Matachewan-Hearst dikes (MHD) represent very large-scale basaltic magmatism that occurred in a stable continental mass (5) at 2.45 Ga (6). The dikes exhibit a north to northwestward trend for a strike distance of over 700 km and an areal extent of approximately 250,000 km2.

The large size of the swarm results in the dikes crossing litho-tectonic boundaries along which Archean blocks (i.e. subprovinces) of markedly different lithologic character have been juxtaposed. The subprovinces can be described as metavolcanic-rich belts, e.g. the Abitibi, Wawa and Wabigoon subprovinces, separated by intervening metasedimentary belts such as the Quetico and English River Gneiss belts (7, 8). A northeast-trending structural discontinuity, the Kapuskasing Structural Zone (KSZ), forms a western boundary along the central portion of the swarm and has been interpreted as an oblique cross section through the Archean crust (9). Studies of KSZ rocks indicate that the Archean-age crust is a complex mixture of mafic to silicic rocks that has been metamorphosed to conditions of amphibolite to granulite facies (5, 10, 11). Geochemical studies indicate that the granulitic portion of the Archean crust in the KSZ has not been depleted to the extent of more 'typical' lower crust elsewhere (10, 11).

Most of the MHD host plagioclase megacrysts whose composition (AN85±5) is identical to the plagioclase in Archean anorthosites of the Superior Province (12, 13). Trace element modeling (14, 15) of a subset of the MHDs suggests that a hierarchy of controls on MHD magma composition occurred. On a local scale, flow differentiation and multiple magma injection produced chemical variability observable within single traverses across dikes. Variations in the concentrations, observed in both intra- and interdike suites, were produced in part by combined replenishment-fractional crystallization (RFC). This process resolves the apparent decoupling of major- and trace element variations (14, 15) and is consistent with the uniform (i.e. non-zoned) plagioclase megacrysts (12). The RFC process, however, is incapable of producing the observed range in values of incompatible element ratios, e.g. $(La/Sm)n$ varies from 0.62 to 2.02. As illustrated in figure 1, the processes of simple fractional crystallization (FC) and RFC, although capable of affecting a significant change in the concentration of a given element, La in this case, are incapable of producing detectable changes in the $(La/Sm)n$ ratio. A pre-RFC stage of evolution, involving other processes, must have operated. Such processes, which are not mutually exclusive, include combined assimilation-fractional crystallization (AFC) and source effects.

In the AFC process, basaltic magma either melts, dissolves or otherwise assimilates crustal material. Energy required for the assimilation is derived from the latent heat of crystallization released as the magma simultaneously crystallizes. The exact trajectory of compositional evolution (e.g. fig. 1) will depend on (a) the

18-4

composition of the parent magma, (b) the composition of
crystallizing phases, (c) the partitioning of elements
between the crystallizing phase and the melt, (d) the
composition of the assimilant and (e) the rate of assimi-
lation to crystallization. Clearly, the AFC process cannot
be described in terms of a simple mix between the parent
magma and the contaminant.

Source effects refer to the impact that chemical
heterogeneity in the mantle source and/or variations in the
melting processes can have on the compositions of the
derivative melts. Numerous studies have shown that the
mantle is non-uniform, both on a local and regional scale.
If different sources are tapped or if a single markedly
heterogeneous mantle is variably melted (16), variations in
incompatible element ratios can be produced (fig. 1).

GEOCHEMICAL MODELS

The MHD magmas inherited a certain amount of
compositional variability, either from the mantle source(s)
from which they were derived, from interaction with lower
crust through AFC, or from both. In this section, these
two alternatives are explored in more detail. Algorithms
of the AFC and melting processes generate hypothetical
magmatic evolution paths. These calculated trajectories are
compared to the actual data to develop and test various
geochemical models. In addition, the calculations help
constrain models regarding the chemical nature of the lower
crust and mantle source(s). This information is pertinent
to arguments of whether or not plate tectonics operated
prior to the generation of the MHD magmas.
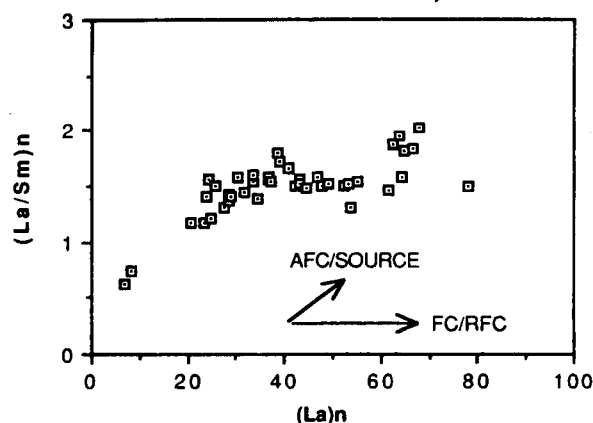
Figure 1.- A plot of chondrite-normalized La against
(La/Sm) for the MHD data (squares). Also shown are
vectors indicating the direction of compositional
evolution via processes discussed in text.

A compositional characteristic of the MHD rocks pertinent to this argument is the relative depletion of the high-field-strength elements (HFSEs), e.g. Ta, Ti, Nb, relative to the rare earth elements (REEs). Values for $(La)n/Ta$ in the MHD range from <40 to >160. Both La and Ta are incompatible elements in basaltic systems; Such a range would not be expected during FC/RFC or melting processes. Fractionation of La from Ta is believed to be characteristic of magmatic rocks, including more silicic differentiates, produced by subduction zone processes (17, 18).

For this study, data for volcanic rocks from a number of tectonic environments was compiled and compared with the MHD. The fields of oceanic island basalt (OIB), ridge basalts (RB) from modern spreading centers, volcanic rocks erupted on continents either as subduction zone related continental arcs (CA) or as continental flood- and rift basalts (CV), and a single example of an oceanic arc, the Finger Bay center in the Aleutians (FB) are plotted in figure 2. The bulk of the MHD plot within the field of CA/CV, distinct from OIB, FB and RB. It would appear that the source of the MHD magmas was not similar to that of modern oceanic islands. Although the MHD lie within a field that includes subduction zone magmas (CA), crustal contamination is possible in all plotted cases. Further, the MHD data define a linear trend from the RB field towards that of crustal rocks (CR).

Assimilation-Fractional Crystallization

Condie et al. (19) argue that crustal contamination was not important in the evolution of the MHD magmas. This conclusion is based on two assumptions that, on close inspection, do not appear valid. They assume that 1) parental magmas of the MHD were identical to modern MORB, and 2) the lower crust in the Superior Province was depleted in LILE, similar to "uplifted Archean granulite terrains". In fact, the more primitive of the MHD magmas have some significant differences with modern MORB, e.g. a low Zr/Nb ratio (<8.6 in contrast to >20 as assumed by these authors). Further, studies of lower crust exposed in the KSZ clearly indicate that these granulite facies rocks are not depleted in a style typical(?) of other such terrains (10, 11). Also implicit in the qualitative arguments of Condie et al. (19) is that contamination is only a two endmember process. As discussed above, it is not. In this study, AFC models were developed from crustal rock data, e.g. a granodiorite xenolith and tonalitic compositions from the KSZ (10). In addition to considering bulk assimilation of silicic granitoids, models reflecting mixing of basaltic magma with partial melts of basaltic precursors ($\pm$
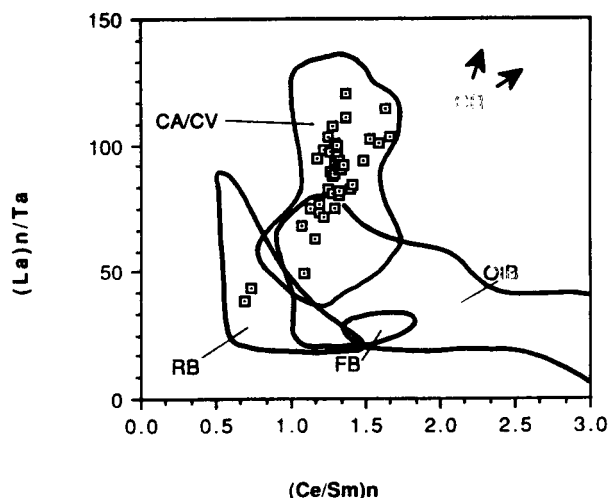
18-6

Figure 2.- A discriminant diagram, plotting $(Ce/Sm)n$ against $(La)n/Ta$ for rocks from the environments of oceanic islands (OIB), spreading centers (RB), an oceanic arc (FB) and rocks emplaced on the continents, either above subduction zones, as flood basalts or in rift zones (CA/CV). The MHD data shown as squares. Arrows point in the direction of crustal rocks (CR).

garnet in the residuum) are considered. These latter assimilants are considered because of the possibility that underplating is an episodic process (4), and that the partial melting of previously underplated basalts might provide a potential contaminant. Experimental results suggest that the major element compositions of such melts will be tonalitic (20). Remelting of the tonalites (partial melts of earlier underplated basalts[?]) would yield liquids in the granodiorite to granite range (21).

In figure 3, the MHD data are shown as a field on the Th-Hf-Ta triangular diagram of Wood (22). The data fall within the field of "destructive plate-margin basalts and differentiates" (22, p. 12). Menzies et al. (23) point out that magmas assimilating Th-rich crustal rocks will be displaced towards the Th apex of the triangle. AFC calculations in this study substantiate this claim. The symbols in figure 3 represent the trajectories of AFC, assuming a depleted parental magma (symbols in field N) and either silicic or mafic granulites as assimilants. Although the various assimilants can not be discriminated, it is clear that the AFC process can explain the MHD data.

The relative abilities of the four assimilants are evaluated in figure 4. Incompatible element ratios are utilized here to essentially filter out the effect of subsequent RFC processes. RFC will not produce significant variations in these ratios (fig. 1). Figure 4 indicates that the MHD data can best be accommodated by AFC processes involving silicic granitoids. The assimilation rate used
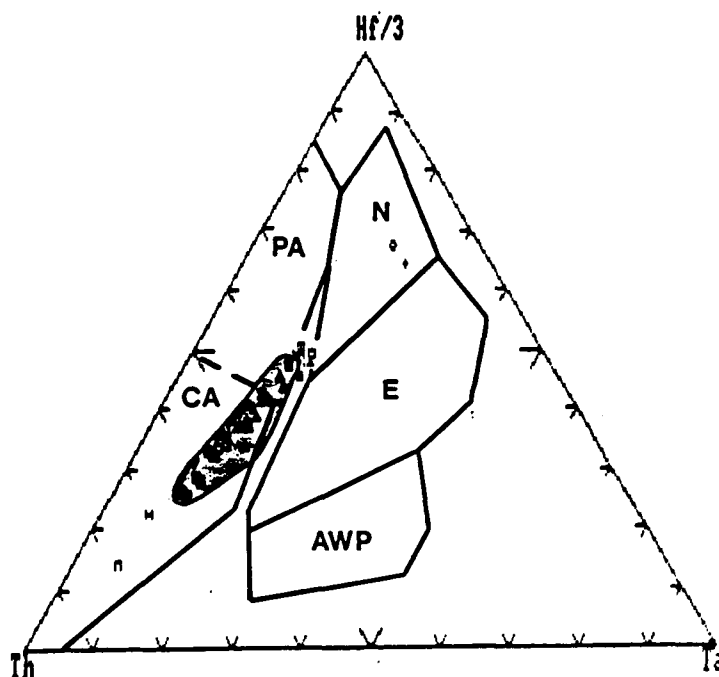
Figure 3.- Triangular Th-Hf-Ta discriminant diagram of Wood (24). Fields are N-type (N) and E-type (E) MORB, alkaline within plate basalts and differentiates (AWP), primitive arc tholeiites (PA) and calc-alkaline basalts and differentiates (CA). MHD data are shown as shaded field and the two points in the N field. Symbols track the trajectories of AFC models.
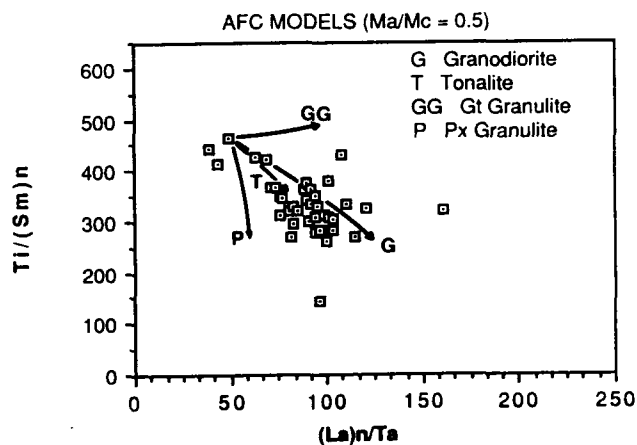

Figure 4.- Plot of (La)n/Ta versus Ti/Sm for MHD data (squares). Vectors represent AFC models (Ma/Mc = 0.5) for various assimilants as labeled.

was 0.5 (= mass of assimilant/mass of cumulates), a rate
consistent with conditions of the lower crust. Several
factors contribute to the scatter of the MHD data. The MHD
data considered in this study occur over an area of 60,000
km2. Given the lithologic diversity of the lower crust
(10, 11), and the possibility of variable assimilation
rates, one can envision a splay of vectors originating from
the parental composition. Heterogeneities within the
mantle source are also possible, although isotopic data
suggest that variations were small and the mantle was
depleted in character (24, 25). The isotopic support for a
depleted source is consistent with the choice of the light-
REE depleted MHD compositions as parental liquids. In
figure 5 the impact of AFC, incorporating small variations
in source compositions, has been evaluated. Compositions
18B and 15B represent a MHD and an older pillowed flow
rock, respectively; both, therefore, reflect variations in
mantle chemistry below the Superior Province. The para-
meters Ta/Ta* and Ti/Ti* (Ta and Ti anomalies) analogous to
Eu anomalies, are calculated from the extended REE diagrams
of Briqueu et al. (18). The AFC calculations portrayed in
this figure indicate that taking into account minor hetero-
geneities in both the mantle source and the lower crust
will account for the bulk of the MHD data.

Further examination of the models and data of figure 5
indicate that some of the MHD compositions appear to
require an assimilant derived from a mafic (i.e. basaltic)
precursor. AFC calculations indicate that this mafic
source rock must have been light-REE enriched; the use of
partial melts of depleted compositions such as 15B and 18B
as contaminants were unsuccessful. Figures 4 and 5 also
indicate that if the mafic rock had a garnet-free mineral-
ogy, partial melts of this assemblage would be unsuccessful
contaminants. The implication of a garnet-bearing residuum
is that the crust must have been thick enough to stabilize
this phase. The requirement for a mixed silicic-mafic
assimilant package emphasizes the heterogeneous nature of
the lower crust (5, 9 and 11).

The relative success of the AFC models is illustrated in
the extended REE diagram given in figure 6. Shown in this
figure is a potential parental composition (15B), a grano-
diorite contaminant (3G), and an evolved MHD composition
(25A). The stippled field is the range of compositions
produced through AFC (Ma/Mc = 0.5) after the original magma
mass has been reduced by 10 to 20 percent. It is evident
that the trace element composition of 25A is consistent
with such a process. Condie et al. (19) argued that signi-
ficant assimilation of silicic crust would change the
silica content of the evolved magma to that uncharacter-
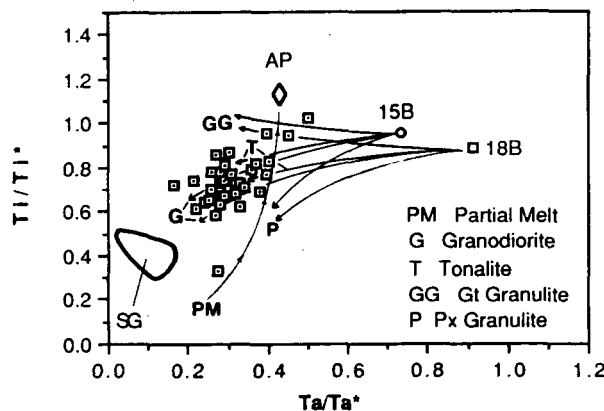istic of basalt. Recalling that the AFC process involves

Figure 5.- Anomalies of Ta versus Ti (see text and 20) for MHD data (squares), various AFC models involving possible parental melts 15B and 18B, and partial melt trajectory from a hypothetical metasomatized amphibole-bearing peridotite (AP). Tonalitic compositions (10) from the KSZ shown as outlined area (SG).

three endmembers (parental magma, assimilant and cumulate phases), it is noted that the liquidus assemblage of the MHD magmas must have been dominated by clinopyroxene and plagioclase (14), both of which have silica contents higher than the parental magma. Fractionation of such phases offset the higher silica content (in the case of 3G, approximately 70 percent) of the assimilant. Qualitative calculations indicate that the AFC model portrayed in figure 6 produces an evolved liquid having a SiO_2 content consistent with that of 25A.

Enriched Mantle Source(s)

In the previous section it was demonstrated that AFC processes can produce the compositional characteristics of the MHDs. In this section, the possibility that these characteristics could alternatively reflect an enriched source is evaluated. Of particular interest is whether or not a source, previously enriched by subduction zone processes, is indicated. The depletion of HFSE relative to the LILE and REE, observed in the MHD data, is interpreted by some to be unique to subduction related volcanism (e.g. 18, 27), resulting from the transfer of the LILE and REE from the slab as a result of dewatering processes (17). The fractionation of LILE and REE from the HFSE, however, may
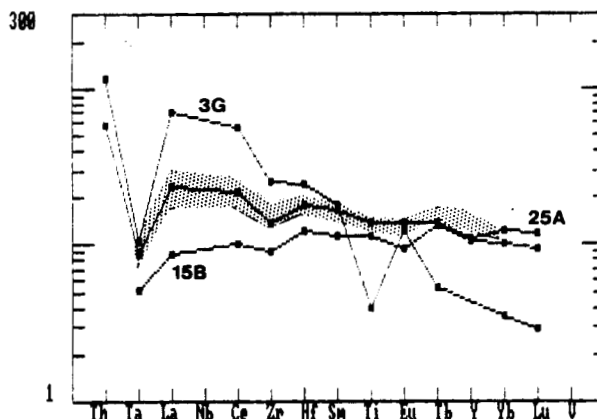
Figure 6.- Extended REE diagram for samples and models
of the MHD. All compositions normalized to values
given by Briqueu et al. (20). Sample 15B represents a
possible parental MHD composition, 3G is a granodiorite
xenolith considered as potential crustal contaminant,
and 25A is intermediate MHD composition. Stippled area
represents range of compositions produced by AFC (Ma/Mc
= 0.5; 15B as parent and 3G as contaminant), with F =
0.9 to 0.8.

also be a function of restite mineralogy, in particular Ti-
rich phases such as sphene or magnetite (18). Further,
Arculus (28) has argued that the LILE enrichment relative
to HFSE may not be restricted to subduction zones, but may
also occur in continental terrains (see figure 2), being
more a function of process than tectonic environment. In
agreement, the partial melt calculations discussed above
indicate that HFSE anomalies can be produced during the
underplating process by the melting of basaltic precursors.
    In order to evaluate whether or not the MHD magmas may
have been derived through variable melting of mantle,
previously enriched above a subduction zone, partial melt
calculations were performed on a hypothetical amphibole-
bearing peridotite. Metasomatic fluids would likely
produce modal amphibole at the expense of clinopyroxene
(29). Progressive melting of such a source (Fig. 5) would
result in decreases in both Ti and Ta anomalies, with Ti
being affected to the greatest extent. Few of the MHD data
conform to such a trend. In figure 7a, the ability of
amphibole to fractionate middle-REEs (Tb) from heavy-REEs
(Yb) is noted. The MHD data clearly do not correspond to

the trend of variable melting of an amphibole-bearing peridotite. Rather, the MHD compositions are subparallel to AFC trajectories involving silicic granitoids as contaminants. Finally, we note that a consistent relation exists between (La)n and Ta/Ta* and Ti/Ti* (Fig. 7b). Recall that shallow-level RFC processes affect the (La)n value but would have no impact on anomaly values. If variable melting of an enriched source was responsible for the anomaly values, then, in order to maintain the relationship shown in figure 8b, the shallow RFC would have to be related to the extent of mantle melting, an improbable relationship.
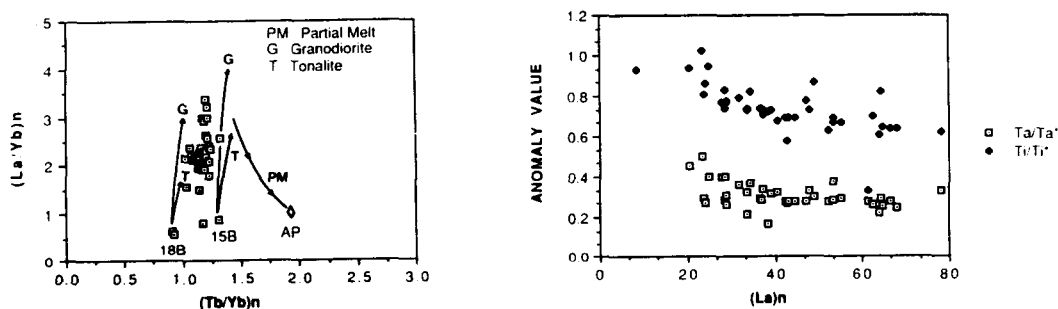
Figure 7.- (a) Plot of (Tb/Yb)n against (La/Yb)n, showing partial melt and AFC trajectories. Symbols as in figure 5. (b) Anomalies of Ta and Ti plotted against (La)n for the MHD data.

CONCLUSIONS

The simplest model for the MHD magmas is AFC, presumably occurring at the base of the crust during underplating. Subduction zone enriched mantle sources are not required. Trace elements suggest that the mantle sources for the MHD were depleted, but possessed a degree of heterogeneity. Rates of assimilation were approximately 0.5 (= Ma/Mc); the contaminant mass was <20%. The contaminant was dominated by tonalites-granodiorites, similar to xenoliths and rocks in the KSZ. Assimilation of partial melts of light-REE and garnet-bearing basaltic precursors may have produced some the MHD magmas. Apparently, previous underplating-AFC processes had already produced a thick crust. The silicic granitoid assimilant for the MHD magmas was probably produced by earlier processing of underplated mafic crust (4, 5, 10, 21 and 30). Calculations suggest that the derived silicic rocks possess negative Ta and Ti anomalies even though they were not the product of subduction.

REFERENCES

1. Basaltic Volcanism Study Project: Basaltic Volcanism on
the Terrestrial Planets.   Pergamon Press, 1981, 1286 pp.
2. Etheridge, M.A.; Rutland, R.W.R.; and Wyborn, L.A.I.:
Orogenesis and Tectonic Process in the Early to Middle
Proterozoic of Northern Australia.   Proterozoic
Lithosphere Evolution, A. Kroner, ed., Am. Geophys. Union
Geodynamics Series Volume 17, 1987, pp. 131-148.
3. Allegre, C.J.: Chemical Geodynamics. Tectonophysics, v.
81, 1982, pp. 109-132.
4. Wass, S.Y.; and Hollis, J.D.: Crustal Growth in South-
Eastern Australia-Evidence from Lower Crustal Eclogite and
Granulitic Xenoliths.   J. Metamorphic Geol., v. 1, 1983,
pp. 25-45.
5. Percival, J.A.; and Card, K.D.: Structure and Evolution
of Archean Crust in Central Superior Province, Canada.
Evolution of Archean Supracrustal Sequences, L.D. Ayres,
P.C. Thurston, K.D. Card and W. Weber, eds., Geol. Assoc.
Canada Spec. Paper 28, 1985, pp. 179-192.
6. Heaman, L.: Geol. Assoc. Canada, Program with Abstracts,
v. 13, 1988.
7. Card, K.D.: Geology and Tectonics of the Archean
Superior Province, Canadian Shield. LPI Tech. Rept., No.
86-04, 1986, pp. 27-29.
8. Percival, J.A.: Metamorphism and Plutonism in the
Quetico Belt, Superior Province, N.W. Ontario. LPI Tech.
Rept. No. 86-04, 1986, pp. 84-85.
9. Percival, J.A.; and Card, K.D.: Archean Crust as
Revealed in the Kapuskasing Uplift, Superior Province,
Canada. Geology, v. 11, 1983, pp. 323-326.
10. Rudnick, R.L.; and Taylor, S.R.: Geochemical
Constraints on the Origin of Archaean Tonalitic-
Trondhjemitic Rocks and Implications for Lower Crustal
Composition. The Nature of the Lower Continental Crust,
J.B. Dawson, D.A. Carswell, J. Hall, and K.H. Wedepohl,
eds., Geol. Soc. Spec. Pub. 24, 1986, pp. 179-192.
11. Ashwal, L.D.; Morgan, P.; Kelley, S.A.; and Percival,
J.A.: Heat Production in an Archean Crustal Profile and
Implications for Heat Flow and Mobilization of Heat-
Producing Elements. Earth Planet. Sci. Lett., v. 85, 1987,
pp. 439-450.
12. Phinney, W.C.; Morrison, D.A.; and Maczuga, D.E.:
Anorthosites: An Analogue Study. Lun. Sci. Conf. XVIII,
1987, pp. 774-776.
13. Morrison, D.A.; Phinney, W.C.; and Maczuga, D.E.:
Archean Anorthosites: Constraints on the Accumulation
Process. Lun. Sci. Conf. XVIII, 1987, pp. 670-671.

14. Nelson, D.O.: Petrogenesis of Basalts from the Archean Matachewan Dike Swarm, Superior Province of Canada. Final Report, NASA/ASEE Sum. Fac. Fellow. Prog., NGT 44-001-800, 1987, pp. 22-1 - 22-20.

15. Nelson, D.O.; Morrison, D.A.; Phinney, W.C.; and Maczuga, D.E.: Geochemistry of Archean Matachewan Dikes of the Superior Province of Canada. EOS, v. 68, 1987, p. 1518.

16. Nelson, D.O.; and Nelson, K.L.: Geochemical Comparison of Alkaline Volcanism in Oceanic and Continental Settings: Clarion Island Versus the Eastern Trans-Pecos Magmatic Province. Mantle Metasomatism and Alkaline Magmatism, E.M. Morris and J.D. Pasteris, eds., Geol. Soc. America Spec. Paper 215, 1987, pp. 317-334.

17. Tarney, J.; Wood, D.A.; Saunders, A.D.; Cann, J.R.; and Varet, J.: Nature of Mantle Heterogeneity in the North Atlantic: Evidence from Deep Sea Drilling. Phil. Trans. R. Soc. Lond. A 297, 1980, pp. 179-202.

18. Briqueu, L.; Bougault, H.; and Joron, J.L.: Quantification of Nb, Ta, Ti and V Anomalies in Magmas Associated with Subduction Zones: Petrogenetic Implications. Earth Planet. Sci. Lett., v. 68, 1984, pp. 297-308.

19. Condie, K.C.; Bobrow, D.J.; and Card, K.D.: Geochemistry of Precambrian Mafic Dykes from the Southern Superior Province of the Canadian Shield. Mafic Dyke Swarms, H.C. Halls and W.F. Fahrig, eds., Geol. Assoc. Canada, Spec. Paper 34, 1987, pp. 95-108.

20. James Beard, unpublished experimental data, 1988.

21. Glikson, A.Y.; Jahn, B.M.: REE and LIL Elements, Eastern Kaapvaal Shield, South Africa: Evidence of Crustal Evolution by 3-stage Melting. Evolution of Archean Supracrustal Sequences, L.D. Ayres, P.C. Thurston, K.D. Card and W. Weber, eds., Geol. Assoc. Canada Spec. Paper 28, 1985, pp. 303-324.

22. Wood, D.A.: The Application of a Th-Hf-Ta Diagram to Problems of Tectonomagmatic Classification and to Establishing the Nature of Crustal Contamination of Basaltic Lavas of the British Tertiary Volcanic Province. Earth Planet. Sci. Lett., v. 50, 1980, pp. 11-30.

23. Menzies, M.A.; Leeman, W.P.; and Hawkesworth, C.J.: Geochemical and Isotopic Evidence for the Origin of Continental Flood Basalts with Particular Reference to the Snake River Plain Idaho, U.S.A. Phil. Trans. R. Soc. Lond. A. v. 310, 1984, pp. 643-660.

24. Shirey, S.B.; and Hanson, G.N.: Mantle Heterogeneity and Crustal Recycling in Archean Granite-Greenstone Belts: Evidence from Nd Isotopes and Trace Elements in the Rainy Lake Area, Superior Province, Ontario, Canada. Geochim. Cosmochim. Acta, v. 50, 1986, pp. 2631-2651.

25.   Machado, N.; Brooks, C.; and Hart, S.R.: Determination of Initial 87Sr/86Sr and 143Nd/144Nd in Primary Minerals from Mafic and Ultramafic Rocks: Experimental Procedure and Implications for the Isotopic Characteristics of the Archean Mantle under the Abitibi Greenstone Belt, Canada. Geochim. Cosmochim. Acta, v. 50, 1986, pp. 2335-2348.
26.   Cattel,   A.: Enriched Komatiitic Basalts from Newton Township, Ontario: Their Genesis by Crustal Contamination of Depleted Komatiite Magma. Geol. Mag. v. 124, 1987, pp. 303-309.
27.   Ellam,   R.M.;   and Hawkesworth,   C.J.:   Elemental and Isotopic Variations in Subduction Related Basalts: Evidence for a Three Component Model. Contrib. Mineral. Petrol., v. 98, 1988, pp. 72-80.
28.   Arculus,   R.J.:   The Significance of Source Versus Process in the Tectonic Controls of Magma Genesis. J. Vol. Geotherm. Res., v. 32, 1987, pp. 1-12.
29.   Neal,   C.R.: The Origin and Composition of Metasomatic Fluids and Amphiboles Beneath Malaita, Solomon Islands. J. Petrol. v. 29, 1988, pp. 149-179.
30.   Gariepy,   C.; and Allegre, C.J.: The Lead Geochemistry and Geochronology of Late-Kinematic Intrusives from the Abitibi Greenstone Belt, and the Implications for Late Archaean Crustal Evolution. Geochim. Cosmochim. Acta, v. 49, 1985, pp. 2371-2383.

LITHOSPHERIC BUCKLING AND INTRA-ARC STRESSES:
A MECHANISM FOR ARC SEGMENTATION

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared By: Kerri L. Nelson

Academic Rank: Lecturer

University and Department: Sul Ross State University
                                          Analytical Laboratory
                                          Alpine, TX   79832

NASA/JSC

Directorate: Space and Life Sciences

Division: Solar System Exploration

Branch: Space Shuttle Earth
                                          Observation Office

JSC Colleague: Charles A. Wood, PhD.

Date Submitted: August 19, 1988

Contract Number: NGT 44-005-803

ABSTRACT

Modern arcs have been subdivided into distinct segments based on features such as changes in the alignment of arc volcanoes, variations in seismic parameters (e.g. frequency, magnitude and distribution of earthquakes) and/or variations in arc lava chemistry. Models for segmentation are varied and include subduction of discontinuites on the ocean floor (transforms; ridges) and weaknesses in the overriding plate. However, these models are not generally applicable between or even within a given arc. Our investigation of 9 ocean-continent convergent margins has shown, however, that there are consistent patterns to segment development. Segments tend to form in 100-300 km lengths regardless of tectonic parameters such as convergence rate or subduction of ocean-floor features. Further, the are consistent relationships between the distribution of calderas in an arc and segment configuration: calderas are concentrated at segment margins and decrease in both size and number towards segment interiors. We model arc segmentation in terms of intra-arc stresses caused by buckling of the downgoing slab as it is subducted. Deformation and/or tearing of the subducted lithosphere induced by buckling provides sites for upwelling of asthenospheric material. As the ashenospheric material underplates the overriding slab, crustal thinning and extension occur, favoring the emplacement of caldera-forming magma chambers. Additionally, the large volumes and short residence times of magmas produced at these points precludes extensive modification of their chemistry by petrogenetic processes (e.g. fractionation, replenishment, assimilation and/or mixing) that are believed to be important in the evolution of more normal "arc-type" calcalkaline compositions such as those found at segment interiors.

INTRODUCTION

Volcanic arcs are often broken into discrete segments which are recognized by 1) changes in the volcanic front, 2) seismological data and/or 3) variations in chemistry along the arc (Stoiber and Carr, 1973; Issacks and Barazangi, 1977; Hughes et al., 1980). Although the presence of segments is well documented in individual arcs, little work has been done in the systematic comparison of segments among a number of arcs to determine what process(es) is (are) controlling the formation of segments within an arc. Factors such as discontinuites in the subducted slab, subduction of aseismic ridges, and/or crustal weaknesses have been invoked by a number of individuals to explain segmentation of given arcs (e.g. Issacks and Barazangi, 1977; Carr, 1984; Kay et al., 1982; Hall and Wood, 1985). To better delineate what factors are involved in segment development, we have evaluated segments from a number of arcs, using both a tectonic and volcanic perspective.

INVESTIGATIONS

Using data from the literature, we compiled a list of pertinent tectonic and volcanic paramters for each arc. The majority of the tectonic data was obtained from Jarrad's (1986) comparison of stresses at convergent margins. To minimize the number of variables influencing segment development, we chose arcs which had been active for at least 10 ma and were characterized by an oceanic plate being subducted beneath a continental plate. Arcs that satisfied these requirements were the Aegean, Ryuku, Central America, the Alaska Peninsula, Cascades, Japan, Kamchatka, Mainland Alaska, and the Central Andes. We were particularly interested in the relationship between stratovolcanoes and calderas within a segment as these volcanic features represent sites of focused magmatism and form in response to differing tectonic stresses (Hildreth, 1981; Nelson and Wood, 1987).

Using the volcanic data base, we noted the size and location of each caldera and stratovolcano for each arc. After plotting the volcanic data on maps, we determined the volcano/caldera distribution, and located the individual segments within an arc based solely on changes in strike of the volcanic front. As mentioned earlier, several criteria have been used to identify arc segments. While use of seismology and chemical data may be useful in refining segment boundaries in arcs, Issacks and Barazangi (1977) noted that the alignment of volcanoes within an arc best defines the individual segments. Further, the alignment of volcanoes is preserved long after subduction ceases, thus providing information on segmentation on

inactive and ancient arcs. This method may also be used where chemical data is insufficient to document geochemically anomalous volcanoes and calderas.

After establishing segments within each arc, we concentrated on isolating factors which may control segment location/formation. It became clear that there was no consistent relationship between location of segment boundaries and subduction of seafloor features. In fact, at the location of segment boundaries, there were no ocean floor features currently being subducted which could be called upon to cause segmentation. This does not preclude previous subduction of ridges, seamounts, transforms etc. which may have initiated earlier segmentation of an arc. However, subduction of such features has been called upon to alternatively focus (Hall and Wood, 1985), terminate (e.g. Issacks and Barazangi, 1977) or have no affect (e.g. Fisher et al., 1981) on volcanism and/or segmentation in an arc. Therefore, while subduction of seafloor features may be important in facilitating segmentation of some arcs, we do not believe it is a prerequisite to segmentation of all arcs.

Tectonic parameters which relate the interrelationship between both the subducting and overriding plate such as convergence rate and strain class were initially considered to be important in controlling the segmentation history of an arc. Strain class (SC) is a quantitative measure of the amount of extension or compression an arc is undergoing, with larger values for SC indicating a more strongly compressive arc. Intuitively, it would seem that the larger the magnitude of interaction between the two plates (e.g. faster convergence, higher SC values) the greater the number of segments that should form. This relationship is clearly not supported by our data set. In fact, the poor correlations observed for the number of segments vs. strain class ($r^2 = 0.344$; Figure 1) and convergence rate
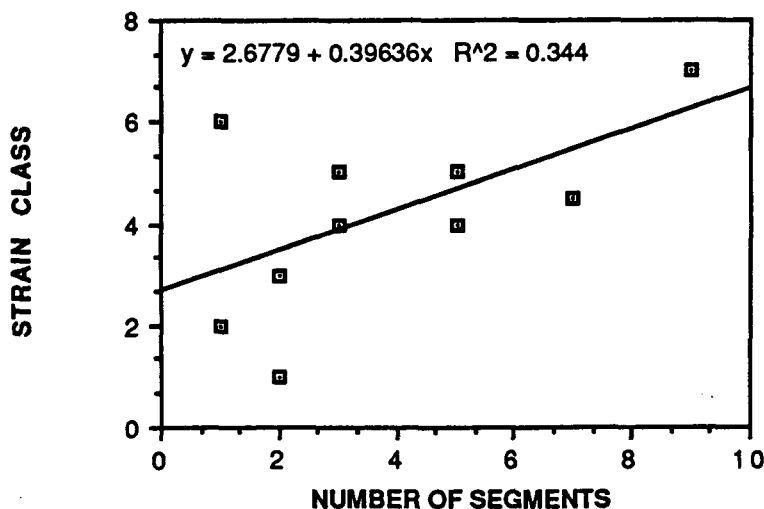

Figure 1.- The variation between number of segments and strain class.factors

19-4

(r^2 = 0.048; Figure 2) would suggest that plate interactions are not primary factors in segmentation. Not only is segment number not controlled by SC,
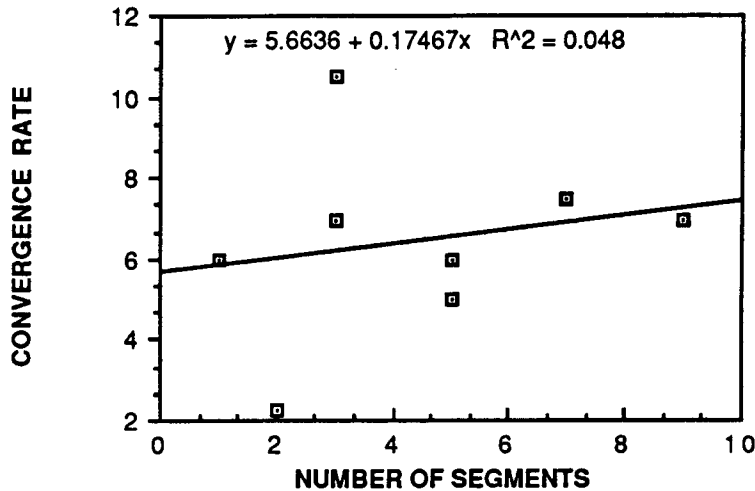

Figure 2.- Variation between segment number and convergence rate.

segment length is also independent of arc stress. Figure 3 illustrates that there is no systematic change in segment length with SC. Although SC 2 is unusual as it is represented by only one arc (Ryuku) which has only one defined segment, there is no clear relationship between SC and segment length: more extensional arcs do not form longer or shorter segments than compressional arcs. Most segments tend to form in lengths between 100 and 300 km (figure 4). This is the same range
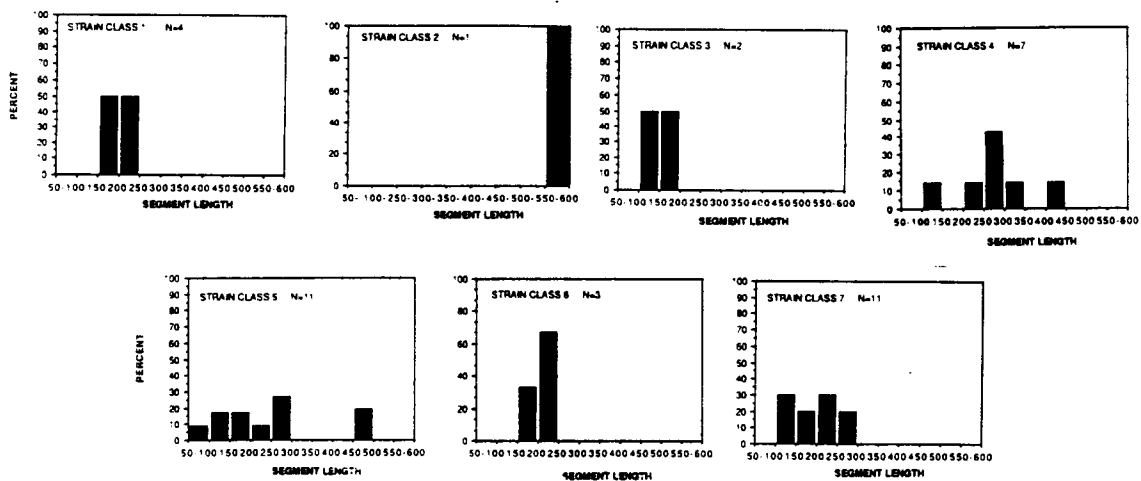
Figure 3.- Distribution of segment lengths as a function of strain class.

19-5
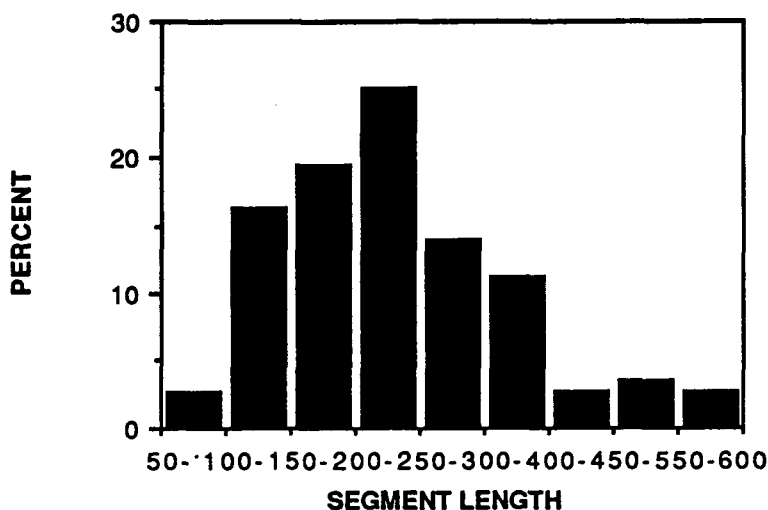
Figure 4.- Frequency of segment lengths.

that has been noted for individual arcs such as Middle America(Stoiber and Carr, 1973) and the Eastern Aleutians (Kienle and Swanson, 1983), indicating that segment length is relatively constant among arcs.

Other factors suggested as important in controlling segments are arc age and crustal thickness. Again, however, the relationship between these factors and segment development is not clear. The poor correlation between number of segments and arc age (r^2=0.197; Figure 5) indicates the segmentation history of an arc is not affected by maturity of an arc. In fact, it would seem that segmentation is established early in the history of an arc and is maintained throughout its development. There is a moderate correlation between crustal thickness and segment number (r^2=0.621; Figure 6) which may indicate that thicker crust may be weakened (fractured ?) by subduction and serve to more effectively segment an arc. However, careful examination of this relationship reveals that arcs with an average crustal thickness of 35-40 km can support between 1 and 7 segments. Clearly, crustal thickness is not a sensitive control or monitor of segmentation. Our statistical analysis (both correlation and stepwise regression) of the available data indicate that the dominant control on the number of segments was neither tectonic style nor character of the overriding plate. Rather, arc length (hence lateral extent of the subducted slab) is the dominant factor in controlling the number of segments (Figure 7; r^2= 0.769; Correlation coefficient = 0.877;). This, coupled with the earlier observation that segment
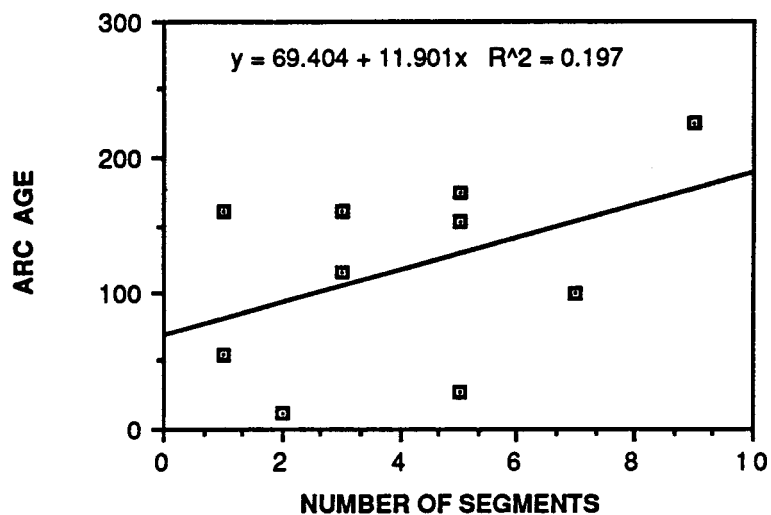
19-6

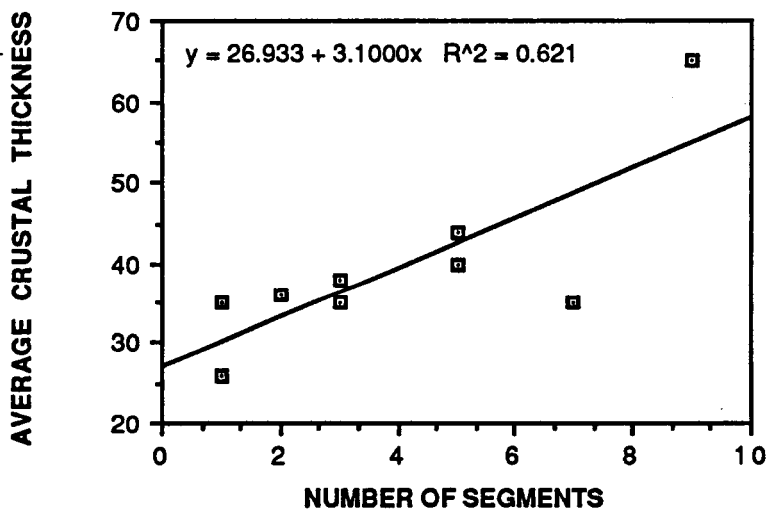**Figure 5.-** Variation between segment number and arc age.

Figure 6.- Variation between segment number and average crustal thickness.
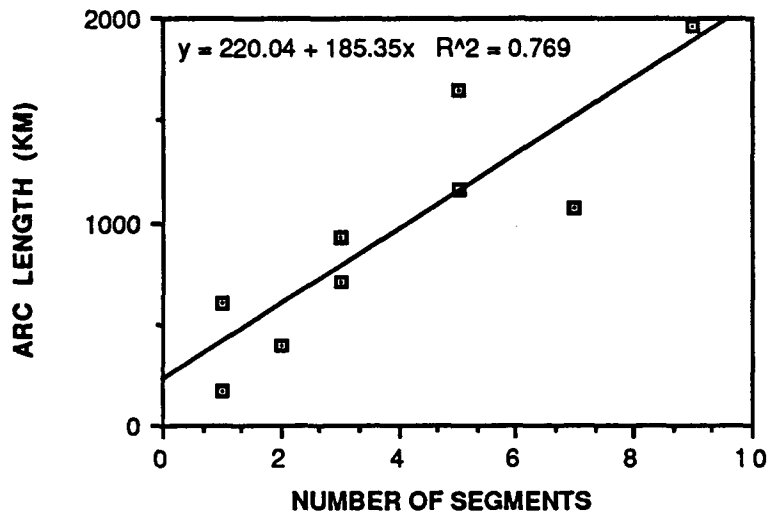
Figure 7.- Relationship between segment number and arc length.

lengths cluster at 100-300 km implies   segmentation is characteristic of arcs and segment number is controlled by the horizontal extent of the downgoing slab.

As large-scale tectonic features   are apparently not important in determining the length and number of segments in an arc, we next evaluated what small-scale (intrasegment) variations in stresses may be recorded in arcs.   Hildreth (1981) noted that stratovolcanoes formed when extension was shallow and/or subordinate while calderas formed   during crustal extension. Previously, we determined that caldera diameter   reflected the amount of extension experienced in an arc: larger calderas formed   as extensional stresses increased (Nelson and Wood, 1987). Therefore, examining   the distribution of   stratovolcanoes and calderas within segments should yield information on   stress variations.   To evaluate   these variations, we divided each segment into percent lengths by effectively "folding" the segment in half: the geographic center of each segment was assigned a value of 50% length, and the percentage length decreased toward segment margins (0%). This allowed for easy comparison of segments of varying lengths. Some consistent relationships between calderas and percent segment length were observed. At segment boundaries, calderas can be located great distances behind the volcanic front while at segment interiors, they   are located primarily on the front (Figure 8a).   Also, the largest diameter calderas are located at segment boundaries and the diameter of calderas decreases toward segment interiors (Figure 8b). Finally, the highest percentage of calderas (38%) are located at segment boundaries and the percentage progressively decreases toward segment interiors (Figure 8c). These consistent relationships, which are seen for all 35 segments, argue for segment

boundaries being more extensional than segment interiors. Further, the relatively smooth decrease in both size and percentage of calderas from segment boundaries to interiors indicates that the change in stress along a segment are gradual and not abrupt as often implied.
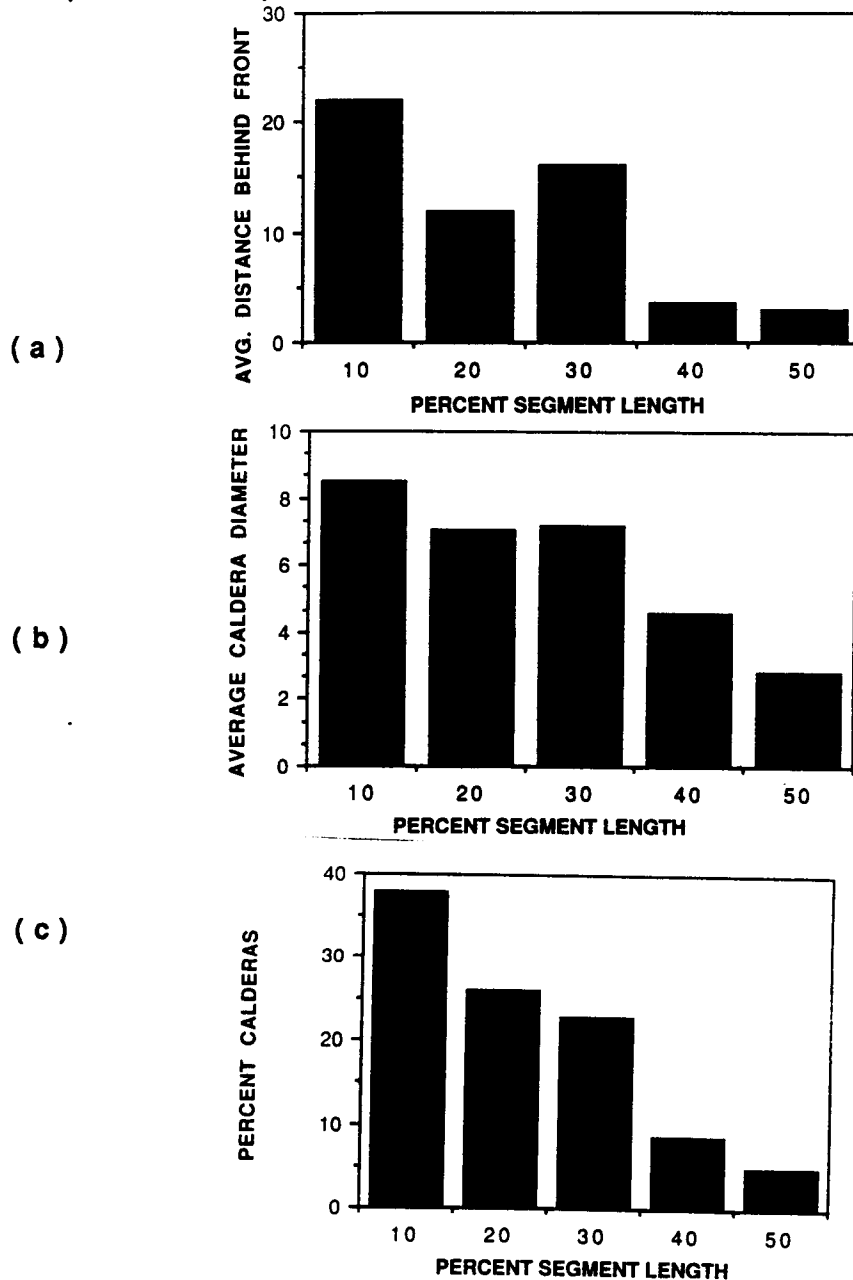
(a)

(b)

(c)

Figure 8.- Distribution of calderas within segments with respect to (a) distance behind the volcanic front; (b) caldera diameter and (c) percentage calderas.

SEGMENTATION MODEL

The similarity of segment development in length, number and stresses among the various arcs investigated is compelling evidence for a common process or processes controlling arc segmentation. We model segmentation as occurring in response to subduction of a spherical oceanic lithosphere beneath an overriding plate . Frank (1968) was first to note that the subduction processes should be influenced by the spherical nature of the subducted lithosphere and Strobach (1973) noted that there was a relationship between the curvature of the trench and the shape of the subducted lithosphere. Bayly (1982) suggested that the arcuate nature of volcanic chains in convergent margins could be produced by buckling of the lithosphere as it was subducted. Experimental modeling of lithospheric buckling reported by Yamaoka et al. (1986; 1987) and Yamaoka (1988) on an arc-arc scale has determined that the subducted lithosphere will deform through processes of buckling, with the wavelength of buckling corresponding to one arc length. Overall, the experimental models predicted the behavior of modern arcs fairly well. Interestingly, however, Yamaoka et al. (1987) noted that, for two arcs, the Aleutians and Middle America, the data achieved a better fit to experimental predictions after being subdivided (segmented?) into separate arcs.

We believe this large-scale (arc-arc) buckling is applicable to, and provides a mechanism for, small-scale (segment-segment) features observed within an arc. As the lithosphere is subducted, it will buckle, creating sites of inflection and/or tearing in the downgoing slab (Figure 9). Once deformed, portions of the slab can act independently of one another, descending at different rates and dips into the mantle with the overlying volcanic chain reflecting the changing geometry of the subducted model. A similar model was suggested by Carr et al. (1982) to account for the transverse boundaries observed in Central America. Further, the points of inflection/tearing in the lithosphere can provide sites for asthenospheric upwelling. Underplating by asthenospheric material at segment margins results in crustal thinning and extension, favoring the emplacement of large caldera-forming magma chambers at relatively shallow levels. The heat source for the crustal thinning/extension would decrease in intensity away from segment margins. Therefore, the distribution of calderas and stratovolcanoes within a segment is a natural consequence of buckling of the subducted slab.

It would seem, then, that lithospheric buckling can explain both the segmented nature of arcs and the distribution of calderas and stratovolcanoes within segments. Additionally, buckling provides a mechanism for explaining the observed chemical variations along segments. Several individuals (e.g. Hughes et al., 1980; Kay et
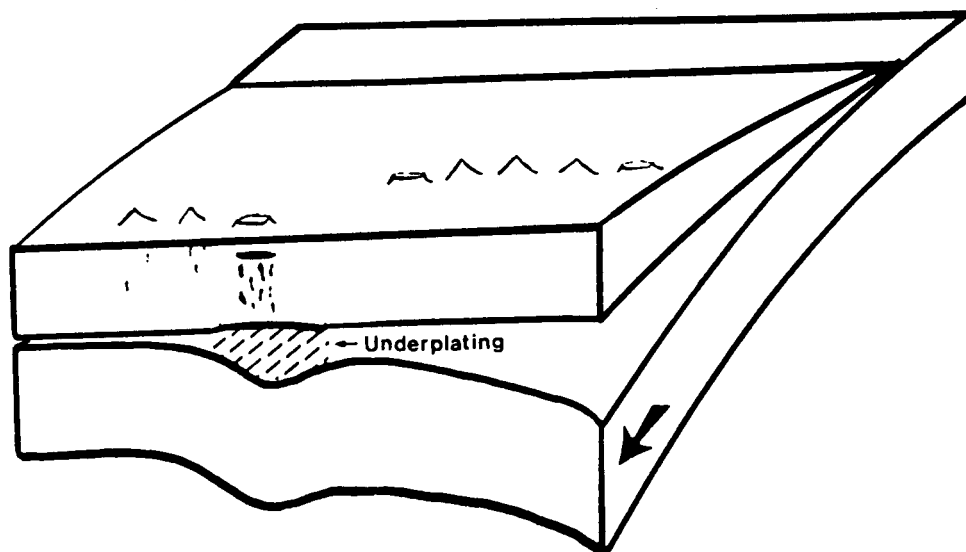
Figure 9.- Schematic model illustrating the relationship between lithospheric buckling and segmentation.

al., 1982; Carr et al., 1982; Wood and Moberger, 1983) have noted the calderas and volcanoes located at segment boundaries erupt material that tends to be chemically atypical for a given arc. Kay et al. (1982) modeled the chemical distribution (calc-alkaline vs. tholeiitic) observed for Aleutian Arc magmas as a function of segment boundary magmas (large tholeiitic centers) rising more quickly through the lithosphere than those magmas (calc-alkaline trend) which were located at segment interiors. Their general model of ascent rate affecting the chemistry of Aleutian magmas has implications for magma evolution at other arcs and can be accommodated within the framework of lithospheric buckling.

As mentioned earlier, points of tearing or inflection produced by lithospheric buckling would create more extensional environments. This is recorded by the presence of large calderas (e.g. Fisher et al., 1981; Carr, 1984; this study) and in the transverse alignment of volcanic cones (e.g. Carr et al., 1982) at segment boundaries. Extensional environments such as segment boundaries, would favor the rapid ascent of magma to shallow crustal levels. In contrast, magmas produced at the more compressional segment interiors would ascend more slowly. The differences in crustal residence times for magmas produced in these two environments is an important control on their chemical evolution.

Magmas produced in compressional environments are envisioned to pool, perhaps several times at different crustal levels, prior to eruption. In these chambers, the magmas will have ample opportunity to undergo processes of fractionation, replenishment, mixing and assimilation. These combined processes will produce the more typical range of arc-type magma chemistry associated with convergent margins. At segment margins, where significant amounts of subcrustal heating occurs in response to underplating of asthenospheric material, more voluminous magmas can be produced. Important in the evolution of these segment boundary magmas is their rapid rise through the crust. Although similar petrologic processes are operative throughout the segment (e.g. fractionation certainly can occur in segment boundary environments), the shorter crustal residence time and larger volumes of segment boundary magmas will place restrictions on the degree of chemical modification and evolution they can experience. As a result, they will tend not to have chemistries typical of arc magmas produced at segment interiors.. Rather than being chemically "anomalous", however, we believe that these segment boundary magmas are reflecting a **normal** tectonomagmatic processes at convergent margins which is produced by lithospheric buckling.

As a final point, Yamaoka et al. (1987) noted that the age and length of the subducted slab controlled the wavelength of lithospheric buckling. Further, Yamaoka (1988) noted that lithospheric bending is "strongly constrained by the lateral (along arc) continuation of the lithosphere". While we observed no correlation between age or length of the subducted slab, and segmentation history, our work on segmentation of modern arcs clearly indicates that the number of segments is controlled by the lateral extent of the subducted slab. As segment length is relatively constant (100-300 km) it would seem that segmentation should occur in all arcs, regardless of discontinuities on the subducting slab and/or crustal weaknesses. Features such as these may affect the distribution of segment lengths about the 100-300 km range and represent a second-order control on segmentation.

CONCLUSIONS

Comparison of segment development of a number of arcs has shown that consistent relationships between segmentation, volcanism and variable stresses exists. We have successfully modeled these relationships using the conceptual model of lithospheric buckling of Yamaoka et al. (1986; 1987). Lithospheric buckling (deformation) provides the needed mechanism to explain segmentation phenomenon: offsets in volcanic fronts, distribution of calderas within segments, variable segment stresses and the chemical diversity seen between segment boundary and segment interior magmas.

REFERENCES

Bayly, B.: Geometry of Subducted Plates and Island Arcs viewed as a buckling problem. Geology, 1982, v. 10, pp. 179-191.

Carr, M.J.: Symmetrical and Segmented Variation of physical and geochemical Characteristics of the Central American Volcanic Front. J. Volcan. Geotherm. Res., v. 20, 1984, pp. 231-252.

Carr, M.J.; Rose, W.I.; and Stoiber, R.E.: Central America. Andesites, R.S. Thorpe, ed., John Wiley and Sons, 1982, pp. 149-166.

Fisher, M.A.; Bruns, T.R.; and Von Huene, R.: Transverse Tectonic Boundaries near Kodiak Island, Alaska. Geol. Soc. America Bull., 1981, v. 92, pp. 10-18.

Frank, F.C.: Curvature of Island Arcs. Nature, 1968, v. 220, p. 363.

Hall, M.L.; and Wood, C.A.: Volcano-Tectonic Segmentation of the Northern Andes. Geology, v. 13, pp. 203-207.

Hildreth, W.: Gradients in Silicic Magma Chambers: Implications for Lithospheric Magmatism. J. Geophys. Res., 1981, v. 86, pp. 10153-10192.

Hughes, J.M.; Stoiber, R.E.; and Carr, M.J.: Segmentation of the Cascade Volcanic Chain. Geology, v. 8, pp. 15-17.

Issacks, B.L.; and Barazangi, M.: Geometry of Benioff Zones: Lateral Segmentation and Downward Bending of the Subducted Lithosphere. Island Arcs, Deep Sea Trenches and Back-Arc Basins, M. Talwani and W.C. Pitman, III, eds., Am. Geophys. Union, Maurice Ewing Series I, 1977, pp. 99-114.

Jarrad, R.D.: Relations Among Subduction Parameters. Rev. of Geophys., 1986, v. 24, pp. 217-284.

Kay, S.M.; Kay, R.W.; and Citron, G.P.: Tectonic Controls on Tholeiitic and Calc-Alkaline Magmatism in the Aleutian Arc. J. Geophys. Res., 1982, v. 89, pp. 4051-4072.

Kienle, J.; and Swanson, S.E.: Volcanism in the Eastern Aleutian Arc: Late Quaternary and Holocene Centers, Tectonic Setting and Petrology. J. Volcan. Geotherm. Res., 1983, v. 17, pp. 393-432.

Nelson, K.L.; and Wood, C.A.: Mid-Tertiary Volcano-Tectonic Development of the Southwestern Cordillera, North America. EOS, 1987, v. 68, p. 1532.

Stoiber, R.E.; and Carr, M.J.: Quaternary Volcanic and Tectonic Segmentation of Central America. Bull. Volcanol.,1973, v. 37, pp. 304-325.

Strobach, K.: Curvature of Island Arcs and Plate Tectonics. J. Geophys., 1973, v. 39, pp. 819-831.

Yamaoka, K.: Spherical Shell Tectonics: on the Buckling of the Lithosphere at Subduction Zones. Tectonophys., 1988, v. 147, pp. 179-181.

Yamaoka, K.; Fukao, Y.; and Kumazawa, M.: Spherical Shell Tectonics and the Evolution of Subduction Zones. Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System, K. Fuchs and C. Froidevaux, eds., 1987, pp. 205-217.

Yamaoka, K.; Fukao, Y.; and Kumazawa, M.: Spherical Shell Tectonics: Effect of Sphericity and inextensibility on the Geometry of the descending lithosphere. Rev. of Geophys.,1986, v. 24, pp. 27-53.

Wood, C.A.; and Moberger, D.: Segment length: A Tectonic Control on Arc Volcanism. EOS, v. 64, p. 845, 1983.

TOWARD AN IMAGE COMPRESSION ALGORITHM
FOR THE HIGH-RESOLUTION ELECTRONIC STILL CAMERA

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1988

Johnson Space Center

Prepared By: Rosalee Nerheim, Ph.D.

Academic Rank: Assistant Professor

University & Department: DePaul University
                                          Department of Computer
                                                  Science
                                          Chicago, IL   60604

NASA/JSC

Directorate: Space and Life Sciences

Division: Man-Systems

Branch: Flight Systems

JSC Colleague: James Ragan

Date Submitted: August 11, 1988

Contract Number: NGT 44-005-803

Abstract

Taking pictures with a camera that uses a digital recording medium instead of film has the advantage of recording and transmitting images without the use of a darkroom or a courier. However, high-resolution images contain an enormous amount of information and strain data-storage systems. Image compression will allow multiple images to be stored in the High-Resolution Electronic Still Camera. The camera is under development at Johnson Space Center. Fidelity of the reproduced image and compression speed are of tantamount importance. Lossless compression algorithms are fast and faithfully reproduce the image, but their compression ratios will be unacceptably low due to noise in the front end of the camera. Future efforts will include exploring methods that will reduce the noise in the image and increase the compression ratio.

Background

A conventional camera uses film to store images. Light enters the lens and strikes the film. A digital camera uses a digital recording medium such as memory chips to store images. It converts the light entering its lens into a voltage corresponding to the intensity of the light. The analog voltage is then converted into a number and the numbers are stored in the camera's memory cartridge. A digital camera has the advantage of transmitting images to a remote location without the necessity of physically sending film. Compared to conventional video, a digital camera has superior image quality and provides flexibility and confidence in image transmission. Because the image is represented by a file of numbers, it can be sent at any transmission rate. Further, error detection and correction methods used to transmit digital files virtually guarantee that the images arrive distortion-free.

Developing such a camera is the goal of the Electronic Still Camera project, which is under the direction of Don Yeates at the Johnson Space Center[1]. It will supplement conventional 35mm cameras on the Space Station. Current plans of the Space Station call for long intervals between shuttle visits. Using a digital camera to transmit images to earth will relieve ground support from waiting 60 to 90 days for high quality pictures.

The digital camera will be contained in a housing similar to a conventional 35mm camera body, and crew members on the Space Station will use it in the same way they would use a conventional 35mm camera. After taking pictures, the crew member will remove the camera's memory cartridge on which the pictures are recorded. To preview the pictures, the crew member will use the Feedback/Downlink Unit. (See figure figure 1.) The same unit will transmit images to earth. The ground station will receive the images,

memory cartridge

camera

onboard monitor

playback/
downlink unit

to earth
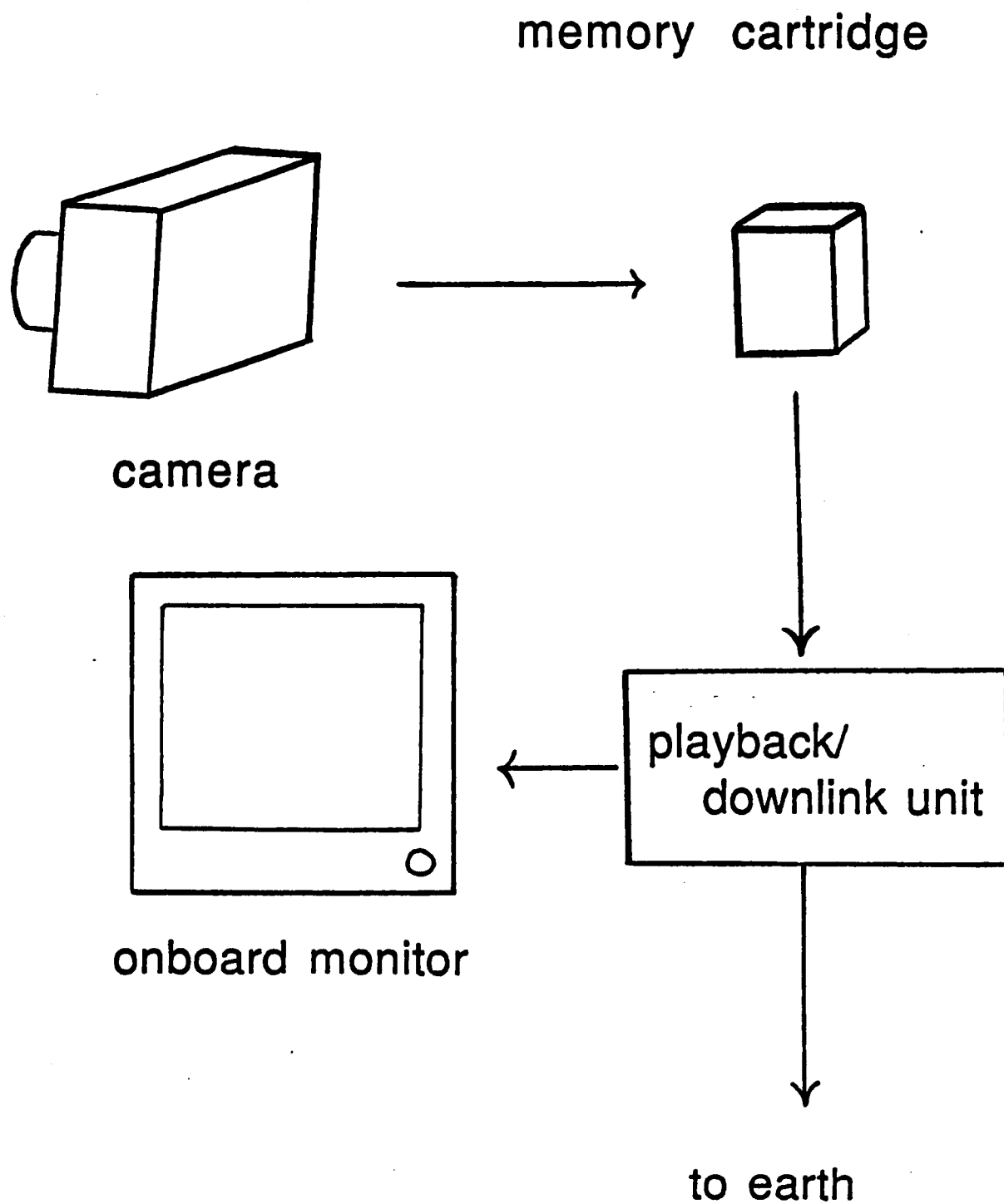
Figure 1.- Electronic Still Camera
Block Diagram

and will have the options of displaying them on a monitor, printing them on film and archiving them for future use.

The amount of memory the camera will need to store an image is tremendous. In order to approach 35mm film quality, the camera's resolution will be 2048x2048 pixels. Proper color response dictates that a pixel have eight bits of information for each of its red, green and blue components for a total of 24 bits. Recording a single image requires 12 Mbytes of memory.

The dimension of the camera body restricts its memory capacity to one image. Even with the one Mbit chips now commercially available, 12 Mbytes of memory requires 108 chips and occupies a volume approximately 4 inches by 4 inches by 6 inches. There is no room inside the camera for additional memory.

A memory cartridge with a one-picture capacity is unsatisfactory. A crew member inside the Space Station would find it irritating to change memory cartridges after every camera shot. For an extravehicular activity, crew members would find this problem extremely troublesome. A solution to the memory capacity problem is to use a compression algorithm in order to store multiple images on the memory cartridge.

Compression is the process of representing an image file in fewer bits[2]. The original image file is compressed into a smaller file. (See figure 2.) Reconstructing the image from the compressed file yields a viewable picture. Important properties of compression algorithms are fidelity, compression ratio and speed. Fidelity is the degree to which the reconstructed image matches the original image. The compression ratio is a ratio of original file size to compressed file size. Speed is either expressed as a proportion relative to a function of the image resolution or in the number of seconds required to compress a file of specified dimension[3,4].
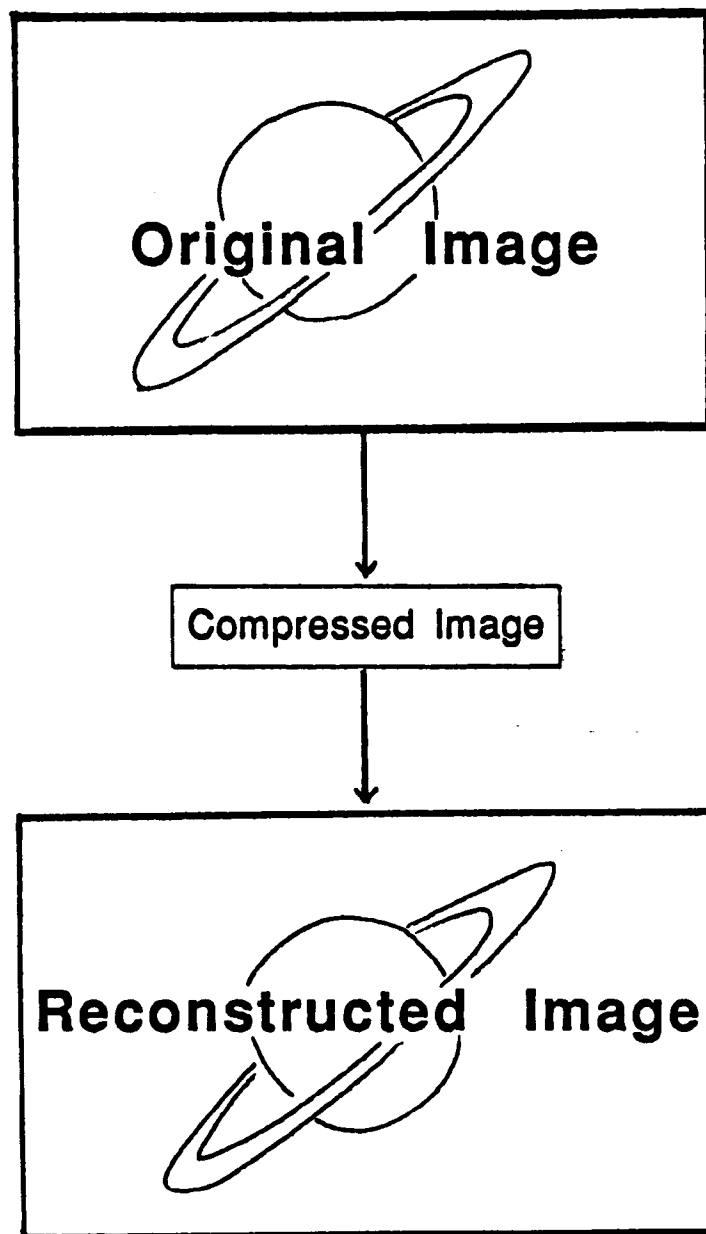
Figure 2.- The Compression Process

Goals

The compression algorithm for the Electronic Still Camera must have high fidelity, a constant compression ratio and high speed. The images from the Space Station must be good enough to be used for public affairs. A constant compression ratio is necessary because the crew member must know in advance how many shots can be stored on a single memory cartridge. The speed is important because the logic in the camera must compress over four million pixels between shots. Waiting more than five seconds between shots would be annoying. For this application, fidelity and speed are more important than a high compression ratio. The compression algorithm will most likely be implemented as a specialized processor in the camera.

Method

I am using all three properties to evaluate the merit of compression algorithms. Objective and subjective measure are necessary to evaluate fidelity. Such objective criteria as root-mean-squared error and signal-to-noise ratio are helpful, but they do not adequately measure fidelity because two pictures with the same rms error may appear to have drastically different visual qualities[5]. A useful subjective method is the pair-comparison method, where observers are shown two images at a time and are asked to give a preference. For measuring speed, I chose to record the number of arithmetic operations per pixel, and the number of separate passes through the image file. The first gives a machine-independent measure that is more precise than O-notation and the second gives a good indication of the amount of parallelism in the algorithm.

The algorithm will be used to compress pictures similar in composition to those taken on the Space Shuttle. I chose nine NASA photographs as being "typical" and used a Howtek scanner to convert them into computer files. The pictures include shots of payloads, the

orbiter, people, equipment and experiments. In addition, I scanned a piece of photoblack paper and a piece of white paper to serve as baseline images. The eleven files serve as test data for the algorithms.

Results

Compression algorithms are called either "lossless" (information-preserving) or "lossy" (entropy-reducing)[6]. Lossless algorithms allow exact duplication of the original image from the compressed form. Lossy algorithms lose some information and an exact duplicate cannot be reconstructed.

I chose to concentrate on the lossless algorithms for the initial study because fidelity is important and because lossless algorithms are generally faster than lossy algorithms. However, the results were disappointing. Figure three is a graph giving the maximum ratios for compression pixel values in the images. For lossless algorithms, there is an upper limit on the compression ratio based on entropy[7]. For a second attack, I compressed the relative differences between adjacent pixels instead of the pixel values. The results were better, but not significantly better. (See figure 4.)

It was troublesome that the image of uniform photoblack did not compress very well. When the image's pixels were magnified on a monitor, I found a lot of noise. For lossless algorithms, high compression ratios occur when long strings of pixels have the same value. Noise corrupts any long string into many shorter strings of slightly differing values. The Howtek scanner used to convert the photographs into computer files uses a charge-coupled device to convert the light into a voltage. The camera will use the same technology, so noise is a problem that must be addressed.
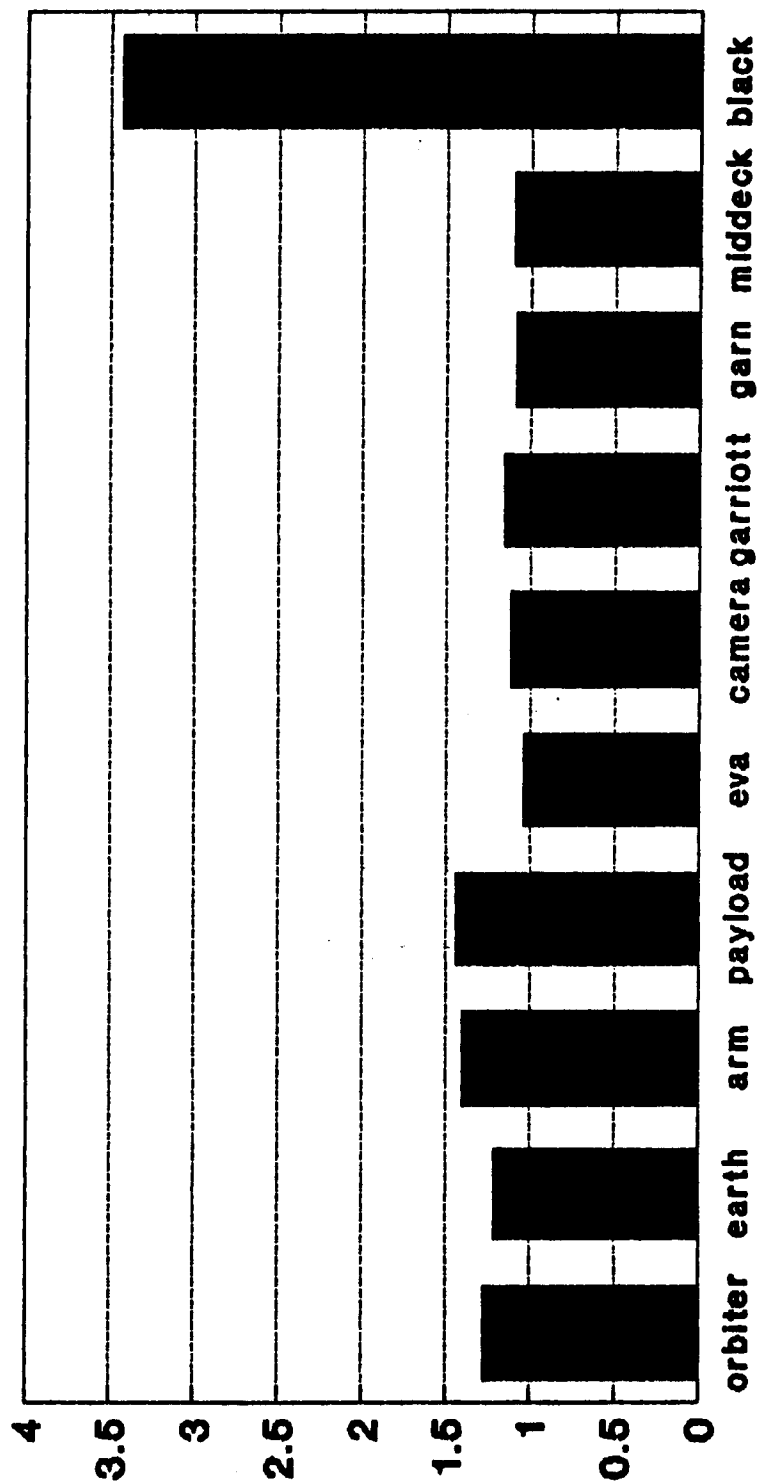
Lossless Compression
Best Case -- Hues

Figure 3.- Compression Ratios for Hues, Best Case
X-axis: picture
Y-axis: compression ratio

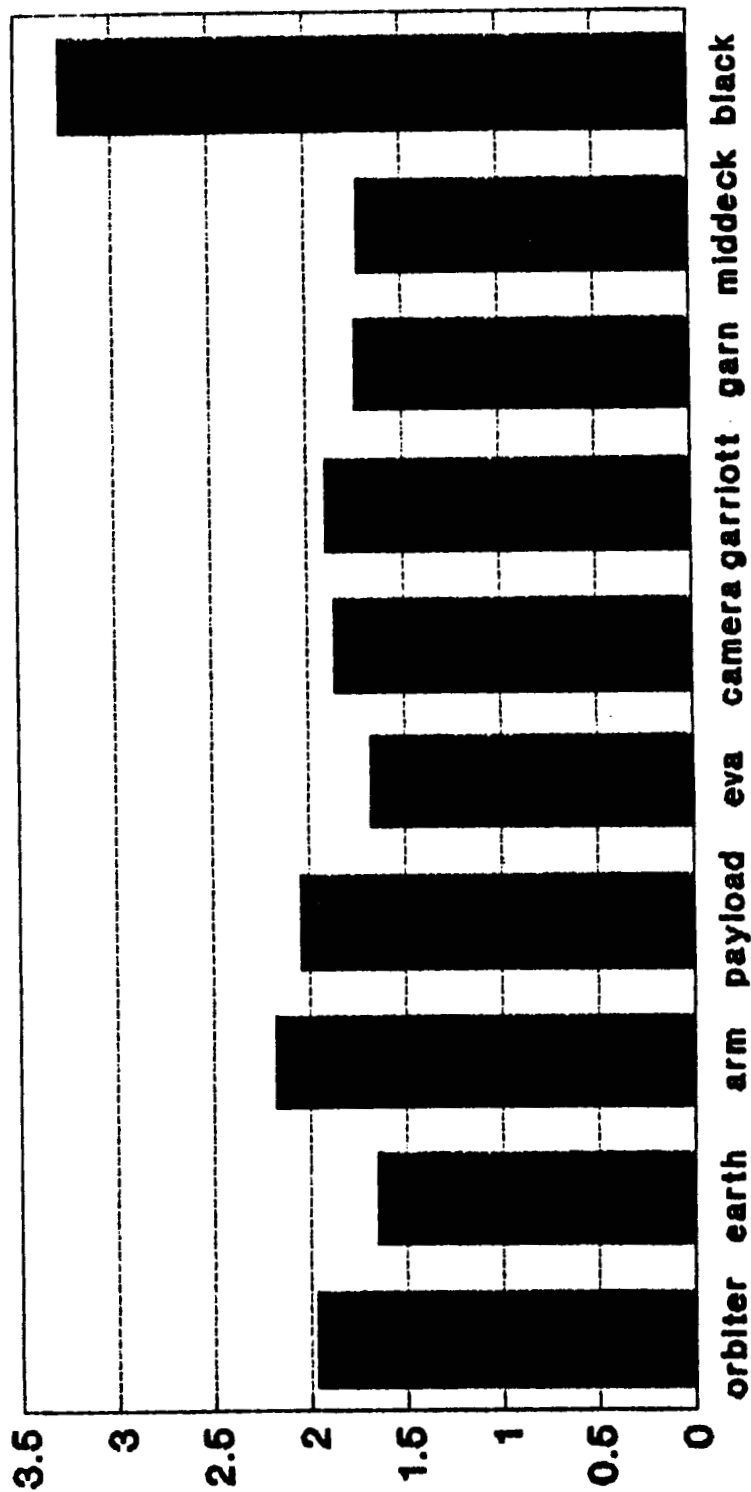# Lossless Compression
Best Case -- Differences

Figure 4.- Compression Ratios for Differences, Best Case
X-axis: pictures
Y-axis: compression ratio

Future Work

One approach to increasing the compression ratio is to "clean up" the image before compressing it. One way to clean up the image is to try calibrating the camera's charge-coupled devices. Another method is to use statistical techniques to compare a pixel with its neighbors and adjust its value if it meets certain requirements.

The act of cleaning up an image loses information. Therefore, I want to explore some lossy algorithms, especially the transform methods, because they reduce noise and compress the image as one process.

References

[1] Don Yeates and Douglas Holland, "High-Resolution Electronic Still Camera (HRESC)", talk given at NASA Johnson Space Center, Man-Systems Division, Flight Equipment Branch, May 13, 1988.

[2] Anil K. Jain, Paul M. Farrelle, and V. Ralph Algazi, "Image Data Compression" pp. 171-226. In Digital Image Processing Techniques, Micheal Eckstrom, ed. New York: Academic Press, 1984.

[3] Azriel Rosenfeld, and Avinash Kak, Digital Picture Processing. New York: Academic Press, 1976. p.125.

[4] Dick Davies, Stevan Eidson and Mike Stauffer, "Improvements in Image Compression." Advanced Imaging (January, 1988) A25-A28.

[5] Rafael Gonzalez, and Paul Wintz, Digital Image Processing. 2nd ed. Reading, Massachusetts: Addison-Wesley, 1987. pp. 256-258.

[6] Murat Kunt, Athanassios Ikonomopoulos, and Michel Kocher, "Second Generation Image Coding Techniques" Proceedings of the IEEE (April 1985) 549-574.

[7]Johannes Moik, Digital Processing of Remotely Sensed Images. NASA SP-431. Washington, DC: NASA Scientific and Technical Information Branch, 1980. pp. 295-296.

EQUATING AN EXPERT SYSTEM TO A CLASSIFIER
IN ORDER TO EVALUATE THE EXPERT SYSTEM

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1988

Johnson Space Center

Prepared By:	Patrick L. Odell
Academic Rank:	Professor
University & Department:	Baylor University Mathematics Waco, Texas 76798

ABSTRACT

A strategy to evaluate an expert system is formulated. The strategy proposed is based on finding an equivalent classifier to an expert system and evaluate that classifier with respect to an optimal classifier, a Bayes classifier.

This paper shows that for the rules considered that an equivalent classifier exists. Also, a brief consideration of meta and meta-meta rules is included. Also, a taxonomy of expert systems is presented and an assertion made that an equivalent classifier exists for each type of expert system in the taxonomy with associated sets of underlying assumptions.

1. INTRODUCTION

It is the purpose of this paper to formulate a formal mathematical relationship between an expert system and a classifier. If indeed the relationship is unique and constructable, then one will be able to evaluate an expert system by comparing that system to a Bayesian classifier which is by definition the best estimate.

In order to formulate a general result one must develop a definition of an expert system and a taxonomy of special types of expert systems. I have selected my glossary from a document prepared for Electric Power Research Institute entitled, Approaches to the Verification and Validation of Expert Systems [1].

The major purpose of this paper is to develop a theoretical foundation for evaluating expert systems.

2. DESCRIPTION OF EXPERT SYSTEMS

A number of definitions of expert systems are found in Artificial Intelligence literature, but the most informative, yet concise one seems to be the following. First, they perform large, tediously complicated and sometimes difficult tasks at expert levels of performance. Second, they emphasize domain specific problem-solving strategies over the more general 'weak methods' of AI. Third, they employ rules of self knowledge to dynamically consider their own inference process and provide explanations or justifications for conclusions reached.

An expert system usually consists of four main parts, as shown in Figure 1.

1. Knowledge Base(s) of structured domain facts and their relationships and heuristics (problem-solving rules),

2. Inference Engine(s) for controlling the application of facts, relationships, and heuristics in solving the problem(s) at hand, usually under the control of an overall Meta-Controller module,

3. Problem Data Base of information about the problem being solved and the history of the solution process, and

4. User Interface providing results/status displays, explanations, and interaction facilities for user inputs.

Either AI or Conventional software techniques can be used for the storage management of the problem data base and the control of the use interface, but AI programming techniques are always used for representing the knowledge base(s) and developing the inference engine(s).

In the sections that follow, the differences between expert systems and conventional software and the types of expert systems, with respect to the ease of verifying and validating them, will be discussed.

2.1 Differences From Conventional Software

Computer scientist claim that expert systems differ from conventional software programs both in the types of problems they solve and their internal structure. Like a human expert, the expert system has to accommodate information that is incomplete, erroneous, or misleading. Yet, a choice of an action or decision among several alternatives must be made. They also claim that a conventional software program works correctly only when inputs are complete, of the proper syntax,

and the problem is ambiguous. These claims are not at all obvious to
others in other discipline who view AI and AI tools as simply computer
language and an approach to helping solve important engineering
problems. However, expert systems and their rule base does pose some
new specific problems in evaluation of software. The problems expert
systems are designed to solve generally fall into the following
categories: interpretation, prediction, diagnosis, design, planning,
monitoring, debugging, repair, instruction and control. Notice, each
of these categories contain a function of observing, recognizing and
may include a call for action.

Expert systems have many structural similarities with conventional
software in the modules which perform conventional tasks such as input
problem data processing, data management, and user interface display
processing. However, the core of expert systems, including the
knowledge bases and inference engines, can be different. Symbolic
representation techniques are primarily used to represent knowledge
versus conventional software's numeric or table-based techniques for
representing information. These symbolic representations include
information about relationships between data items and data item
aggregations and information about heuristics (procedures and rules)
for problem solving, which are not represented in conventional data
bases. Expert systems are developed to solve problems which may not
have easily formalized or algorithmic solutions, so the problem-solving
apparatus (inference engines) must use unconventional methods, in-
cluding heuristic-guided search, symbolic inferencing, generation and
test of solutions, and constraint-based reasoning.

It is these differences from conventional software which make
expert systems interesting and applicable, and yet also less easily
tested, verified and validated. Because their solution methods are
often potentially unbounded and not prescribed in a straightforward
"recipe" format like conventional solution methods, they are therefore
harder to "prove" correct.

2.2 Expert System Types

There are many ways of dividing expert systems into types, by
application, by method of reasoning, etc. Examples of these groupings
are found in Table 2-1 and Figure 2-1. In this discussion, expert
systems will be divided into types relative to the complexity and
difficulty of performing V&V on each type. The types are listed in
Table 2-0.

TABLE 2-0.- EXPERT SYSTEM TYPES

NUMBER	NAME
1	Simple, based on Codified Knowledge
2	Simple with Uncertainty Handling
3	Simple, based on Elicited Knowledge
4	Elicited with Uncertainty Handling
5	Complex
6	Complex with Uncertainty Handling

The first type of expert system, Simple, is developed through the
straightforward encoding of validated and verified decision tables
and/or procedure trees. Its search space is small and examined with
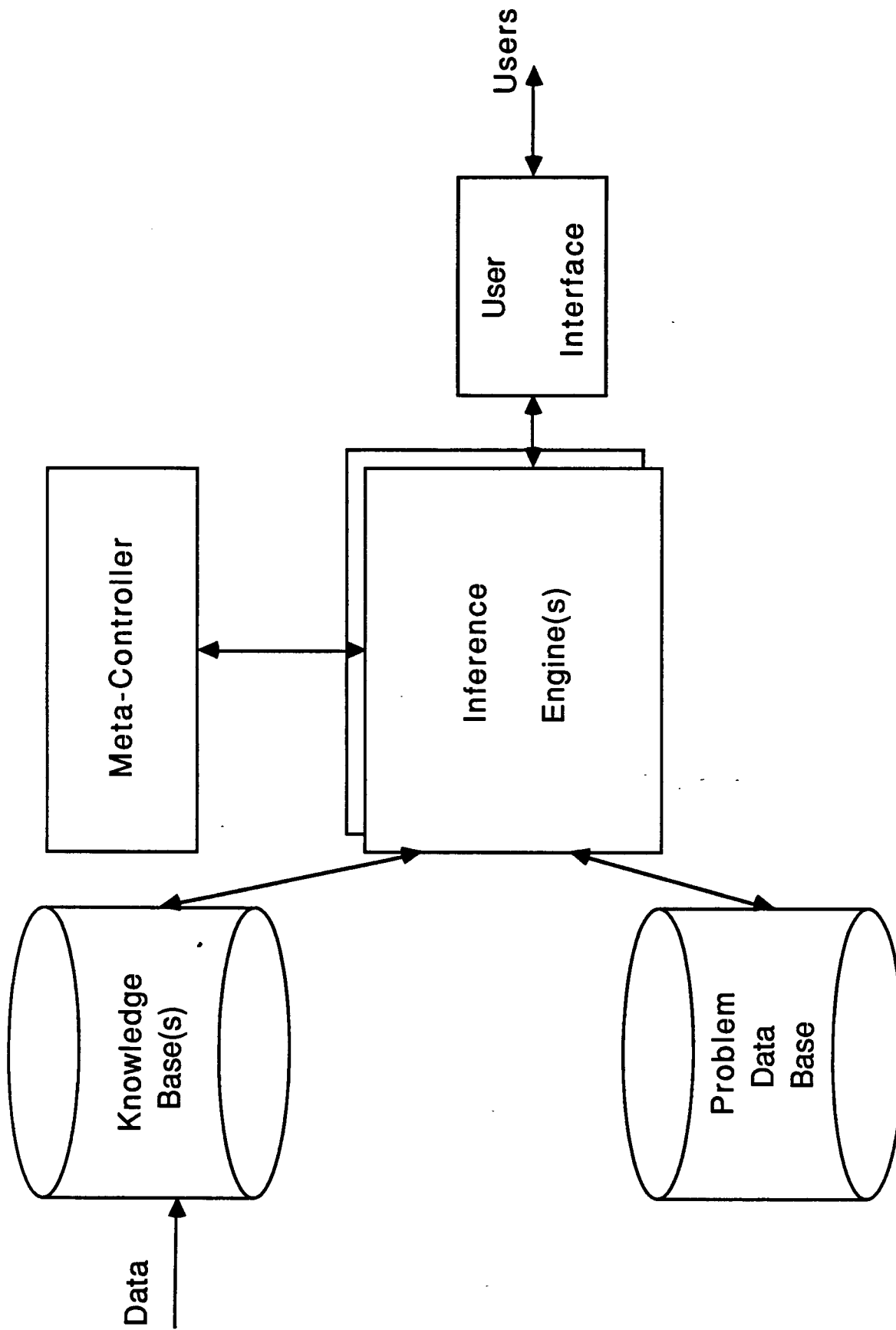exhaustive search techniques or large and factorable and examined with

Figure 1-1. Typical Expert System Components

Table 2-1

GENERIC CATEGORIES OF KNOWLEDGE
ENGINEERING APPLICATIONS

Category	Problem Addressed
Interpretation	Inferring situation descriptions from sensor data
Prediction	Inferring likely consequences of given situations
Diagnosis	Inferring system malfunctions from observables
Design	Configuring objects under constraints
Planning	Designing actions
Monitoring	Comparing observations to plan vulnerabilities
Debugging	Prescribing remedies for malfunctions
Repair	Executing a plan to administer a prescribed remedy
Instruction	Dianosing, debugging, and repairing student behavior
Control	Interpreting, predicting, repairing, and monitoring system behaviors

1

| Small Solution Space |
| Requirements → Data Reliable & Fixed |
| Reliable Knowledge |

| Exhaustive Search |
| Prescriptions → Monotonic Reasoning |
| Single Line of Reasoning |

2

Unreliable Data or Knowledge
Combining Evidence from Multiple Sources
Probability Models
Fuzzy Models
Exact Models

3

Time-Varying Data
State-triggered Expectations

4

Big, Factorable Solution Space
Hierarchical
Generate-and-Test

5

No Evaluator for Partial Solution
Fixed Order of Abstracted Steps

6

No Fixed Sequence of Subproblems
Abstract Search Space

7

Suproblems Interact
Constraint Propogation
Least Commitment

8

Efficient Guessing is Needed
Belief Revision for Plausible Reasoning

9

Single Line of Reasoning Too Weak
Multiple Lines of Reasoning

10

Single Knowledge Source Too Weak
Heterogeneous Models
Opportunistic Scheduling
Variable-Width Search

11

Representation Method Too Inefficient
Tuned Data Structures
Knowledge Compilation
Cognitive Economy

Figure 2-1. Expert System Types Based on Reasoning Methods

21-7

enumerative search techniques, for which proofs exist of the best answer. A single line of reasoning is used and reasoning is monotonic, approaching a single, best answer. This type roughly corresponds to types 1,3, and 4 in Figure 2-1.

The most important feature besides simplicity for this type is that the knowledge to be encoded into the expert system has already been written down in some form and tested for correctness. If the codified decision tables or procedure trees must be augmented with non-proved knowledge, then the expert system does not fall into this type category. So, V&V of this type of expert system involves merely proving that it correctly represents and uses the knowledge, not that the knowledge itself is correct.

The second type of expert system, Simple with Uncertainty Handling, has the features of simplicity and proven knowledge in common with the first, but also includes certainty factors, fuzzy logic, probabilities, or some other method of dealing with uncertain information. This type of expert system roughly corresponds to type 2 in Figure 2-1. There are two general types of uncertainty, uncertainty on the existence or value of knowledge items or their relationships and uncertainty on the rules, reflecting the expert's uncertainty of the applicability of the rule on the antecedent conditions or the appropriateness of the conclusions. Expert systems may incorporate either or both forms of uncertainty. There are various methods of combining certainty factors which may be used alone or in combination in an expert system. The decision of which combination function(s) to use is complex and requires analysis and judgment on the part of the knowledge engineer and expert and effects the reliability of the system.

Even if there exist pre-proven uncertainty values on knowledge items or rules, the method of combining those factors most-likely is not pre-proven and tested. And, usually, the values themselves do not pre-exist and must be elicited from experts. An example is the Probability Risk Assessment (PRA) prior event probabilities. Even though these probabilities pre-exist and have been determined to be correct, to the best knowledge of experts, how the probabilities can be combined and what probabilities to place on rules have not been predetermined and tested. Therefore, uncertainty handling complicates the V&V of simple expert systems by adding the need to V&V the certainty factors and their combination method(s).

The third type of expert system, Simple based on Elicited Knowledge, is probably the most common type in existence. The simplicity of the previous two types is still evident, probably because an expert system building tool or environment was used to develop the system. However, the knowledge (items and rules), or some portion of it, does not pre-exist in a tested form and must be elicited from usually one expert. Successive prototypes are developed and tested to verify that the knowledge required to solve the problem is actually being elicited and that it is encoded properly.

The fourth type of expert system, Elicited with Uncertainty Handling, is a combination of the second and third. Its knowledge has been elicited from an expert and it includes certainty factors. An example of this type of expert system is the Mycin system in the field of medical diagnosis developed by Stanford University. V&V of this type of expert system would include testing the correctness of the knowledge, the correctness of the knowledge implementation, the correctness of the uncertainty factors, and the correctness of the

uncertainty combination functions(s).

The fifth type of expert system, Complex, eliminates all or some of the simplicity assumptions of the previous four, as listed in the description of the first type of expert system. The search space(s) can be infinite or large and unfactorable, subproblems within the problem may interact in various ways, constraint-based reasoning may be used to limit search and multiple lines of reasoning may be pursued to independently produce candidate answers. This type of expert system roughly corresponds to types 5-11 in Figure 2-1. Usually the knowledge is elicited from more than one expert, so conflicting heuristics may arise. Also, usually expert system building tools are not used to implement these systems because the tools do not allow the complexities to be built in.

Examples of the fifth type of expert system include autonomous vehicle control and battle management applications, such as those in the Strategic Computing Program funded by the Defense Advanced Research Projects Agency (DARPA, 1986). Expert systems of this type are just now beginning to work in a primitive form, and V&V of these systems is still a research issue. It is very unlikely that this type of expert system will be developed for non-military application any time in the near future, so V&V of this type of expert system will not be discussed.

The sixth type of expert system, Complex with Uncertainty Handling, adds the complicating factor of uncertainty handling to the fifth, making V&V even more impossible and even more of a research issue. This type also will not be discussed.

3. THE MAIN RESULT

Consider a simple expert system represented by the following rules involving two conditions C_1 and C_2 which can occur or fail to occur and four actions A_0, A_1, A_2, and A_3.

R0   If $C_1 \wedge C_2$, then A_0,

R1   If $C_1 \wedge \overline{C_2}$, then A_1,

R2   If $\overline{C_1} \wedge C_2$, then A_2, and

R3   If $\overline{C_1} \wedge \overline{C_2}$, then A_3,

which can also be represented in vector notation

R0   If   (0, 0), then (0)

R1   If   (0, 1), then (1)

R2   If   (1, 0), then (2)

R3   If   (1, 1), then (3)

Note that the set of rules { R0, R1, R(2), R(3) } can be thought of and modeled by a function R

$$R: \quad (r_1, r_2) \rightarrow \{ 0,1,2,3 \}$$

$$\text{where} \quad r_1 \text{ and } r_2 \in \{ 0,1 \}$$

Suppose further that

$$p_0 = \text{Pr } [C_1 \wedge C_2] \; ,$$

$$p_1 = \text{Pr } [C_1 \wedge \bar{C}_2] \; ,$$

$$p_3 = \text{Pr } [\bar{C}_1 \wedge C_2] \; , \text{ and}$$

$$p_4 = \text{Pr } [\bar{C}_1 \wedge \bar{C}_2] \; ,$$

are the a priori probabilities that the system are in the respective states,

$$S_0 = C_1 \wedge C_2 \; ,$$

$$S_1 = C_1 \wedge \bar{C}_2 \; ,$$

$$S_3 = \bar{C}_1 \wedge C_2 \; , \text{ and}$$

$$S_4 = \bar{C}_1 \wedge \bar{C}_2 \; ,$$

Also let data be taken to make a decision as to which state the system belongs. The data can be vector valued, dependent or independent, known exactly (deterministic) or observable (random). In order to determine if C_1 occurs, we follow the usual logic of statistical classification theory [2]. Here if

$$X = (X_1, X_2, \ldots, X_p)^T \text{ and } Y = (Y_1, Y_2, \ldots, Y_q)^T$$

denotes the observations, then the recognition rules are

RR1   If $X \in R_1$, then $C_1 = 0$.

If $X \notin R_1$, then $C_1 = 1$.

RR2   If $Y \in R_2$, then $C_2 = 0$.

If $Y \notin R_2$, then $C_2 = 1$.

We use the notation $\hat{C}_j$ to denote that we estimate the value of C_j and do not know that value exactly. The classification regions R1 and R2 are selected by the expert or optimally as Bayes regions.

Note that the set of rules when described as above is equivalent to a classical statistical classification problem. We know the solution to this problem; that is, the best classifier is a Bayes classifier. A Bayes classifier is one in which the regions R_1 and R_2 are selected judiciously to assure that the expected costs of misclassification are minimized.

Probabilities that an expert system commits errors can now be computed when the probability density functions $f(x)$ and $f(y)$ of X and Y are known. That is,

$$\text{Pr } [\hat{C}_1 = 0 \mid C_1 = 1] = \int_{R_1} f_1(x) dx$$

$$\text{Pr } [\hat{C}_2 = 0 \mid C_2 = 1] = \int_{R_2} f_1(y) dy$$

and the probability of not making an error is given

$$\Pr[\hat{C}_1 = 0 \mid C_1 = 0] = \int_{R_1} f_0(x)\,dx$$

$$\Pr[\hat{C}_2 = 0 \mid C_2 = 0] = \int_{R_2} f_0(y)\,dx$$

Note that if X and Y are independent then various combined probabilities can be computed easily. If X and Y are not independent then the joint probability density function is required.

One can stack the observations into a new random vector $Z = (x,y)^T$ and establish Z as the data vector. This form allows one to consider the effect of the covariance structure of X and Y.

Suppose one introduce some "beliefs" concerning the rules. Let us model them in the following form.

$$RB_0 \qquad \Pr[R0 \text{ is true}] = q_0$$

$$RB_1 \qquad \Pr[R1 \text{ is true}] = q_1$$

$$RB_2 \qquad \Pr[R2 \text{ is true}] = q_2$$

$$RB_3 \qquad \Pr[R3 \text{ is true}] = q_3$$

This in many applications can affect the action, that is, the actions are modified to

R0 If $C_1 \wedge C_2$, then $\Pr[A_0] = q_0$

and $\Pr[A_0'] = 1-q_0$,

R1 If $C_1 \wedge \bar{C}_2$, then $\Pr[A_1] = q_1$

and $\Pr[A_1'] = 1-q_1$,

R2 If $\bar{C}_1 \wedge C_2$, then $\Pr[A_2] = q_2$

and $\Pr[A_2'] = 1-q_2$, and

R3 If $\bar{C}_1 \wedge \bar{C}_2$, then $P[A_3] = q_3$

and $\Pr[A_3'] = 1-q_3$

These rule can be modelled as stochastic actions which are applied with a specified probability. This formulation introduces the concept of stochastic actions. This in turn generates a requirement for conflict resolution rules for handling uncertainties of this type. To example consider the following two types of conflict resolution rules.

CRR1: If $q > 1-q$, then select action A, if $q \leq 1-q$, select action A'.

CRR2: If the expected cost of A_0 is greater than expected cost of A_0', select A_0'; if not A_0.

CRR3: Generate a uniform random number N on the interval [0,1], if $0 \leq N \leq q$ then select action A_0; if $q < N \leq 1$, then select A_0'.

The first two rules are deterministic rules while CRR3 is a stochastic rule. Clearly, these rules can be extended to more than two actions per rule.

What we have done is given a precise formulation of an expert system which in turn gives a mathematical model for evaluating the expert system as compared with on optimal Bayesian classifier.

Let us now consider the problem of evaluation. Let

$$R_1 = \{X; \underline{X} \leq X \leq \bar{X}\}$$

$$R_2 = \{Y; \underline{Y} \leq Y \leq \bar{Y}\}$$

then we see that directly

Figure 3.1. The (X,Y) plane partitioned into action regions based on an Expert System if

$$X \in R_1 \text{ and } Y \in R_2, \text{ then } A_0$$

$$X \in R_1 \text{ and } Y \notin R_2, \text{ then } A_1$$

$$X \notin R_1 \text{ and } Y \in R_2, \text{ then } A_2$$

$$X \notin R_1 \text{ and } Y \notin R_2, \text{ then } A_3$$

where R_1 and R_2 are the regions defined by the expert.

However, a Bayesian classifier would give different regions R_1 and R_2, the optimal regions. See Figure 3.2.
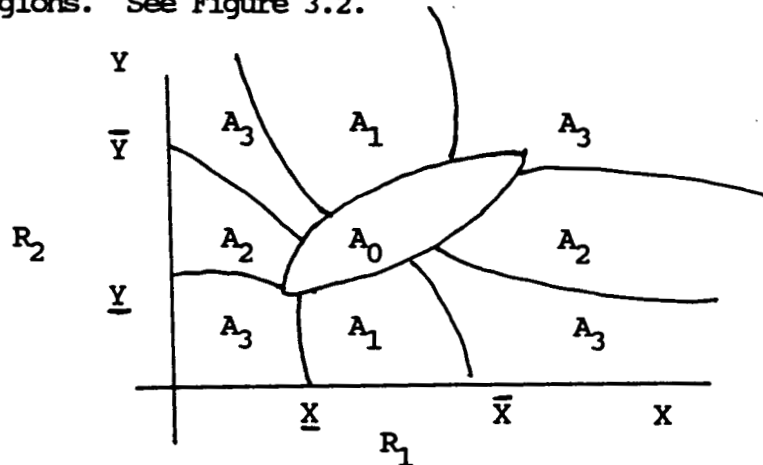
Figure 3.2. The (X,Y) plane partition into action regions based on Bayesian classifier.

Now consider the following meta rule and a meta-meta rule:

MR1: If after K trials

$$\hat{c}_1 \wedge \hat{c}_2 \quad \text{occurs } K_0 \text{ times}$$

$$\hat{c}_1 \wedge \hat{\bar{c}}_2 \quad \text{occurs } K_1 \text{ times}$$

$$\hat{\bar{c}}_1 \wedge \bar{c}_2 \quad \text{occurs } K_2 \text{ times}$$

$$\bar{\hat{c}}_1 \wedge \hat{\bar{c}}_2 \quad \text{occurs } K_3 \text{ times}$$

then replace (a's are preselected known parameters),

$$P_0 \text{ with } a_0 P_0 + (1-a_0) \; K_0/K$$

$$P_1 \text{ with } a_1 P_1 + (1-a_1) \; K_1/K$$

$$P_2 \text{ with } a_2 P_2 + (1-a_2) \; K_2/K$$

$$P_3 \text{ with } a_3 P_3 + (1-a_3) \; K_3/K$$

and compute new regions R_1 and R_2.

MMR1: Let $K(j) = K + j \; 50$.

The argument appears reasonable that an expert system can be modelled mathematically by finding an equivalent classifier and the expert system can then be evaluated by evaluating that classifier.

4. CONCLUDING REMARKS

The types of expert systems listed in Table 2.2 and discussed in Section 2 of this paper may now be defined precisely by defining an associated classifier when different amount and/or kind of knowledge differs.

This allows evaluation of expert systems to be done in a direct way compatible with the theory of classification.

5. REFERENCES

Groundwater, E.H., Donnell, M.L., and Archer, M.A., "Approaches to the Verification and Validation of Expert Systems for Nuclear Power Plants, Electric Power Research Institute, EPRI Np-5239 Final Report, July 1987.

Anderson, T.W., An Introduction to Multvariate Statistical Analysis, John Wiley and Sons, Second Edition (1984) pp. 195-241.

FEASIBILITY OF USING HIGH TEMPERATURE SUPERCONDUCTING
MAGNETS AND CONVENTIONAL MAGNETIC LOOP ANTENNAS
TO ATTRACT OR REPEL OBJECTS AT THE
SPACE STATION

Final Report

NASA/ASEE Summer Faculty Fellowship Program---1988

Johnson Space Center

Prepared By: Manjit S. Randhawa, Ph.D.

Academic Rank: Professor

University & Department: Southern University
Electronics Engineering Technology
Baton Rouge, LA 70813

NASA/JSC

Directorate: Engineering

Division: Tracking & Communications

Branch: Electromagnetic Systems Branch

JSC Colleagues: G. Dickey Arndt, Ph.D.
J. H. Suddath

Date Submitted: August 19, 1988

Contract Number: NGT 44-005-803

ABSTRACT

A study was undertaken to see if magnetic forces can be used at the Space Station to attract or repel spacecrafts such as the Orbital Manuevering Vehicle (OMV) or the Orbiter. A large magnet, in the form of a current loop, is assumed to be placed at the Space Station and another one on the spacecraft. The expression for the force between the two dipoles (loops) is obtained. Using a force of 15 Newtons (3.4 pounds) in order to move the spacecraft, the number of ampere-turn needed in the current loops was calculated at various distances between them.

The expression for the force of attraction between a current loop and a soft magnetic material was also examined and the number of amp-turn needed to provide a force of one-tenth of a pound at various distances is also calculated. This one tenth of a pound force would be used in a life line system for the retrieval of an adrift crewman or tool at the Space Station.

The feasibility of using conventional antenna on the Station and the incoming vehicle for attraction or repulsion was also examined.

INTRODUCTION

In the earth environment, strong magnetic forces are needed at short distances for attraction or repulsion, while in space, weak forces are needed at large distances due to the microgravity environment. The force needed to move an Orbital Manuevering Vehicle (OMV), weighing 20,000 pounds, with an acceleration of 150 μg at low earth orbit will be about 3 pounds, while a force of 3.4 pounds (15 Newtons) will move the Orbiter (weighting 225,000 pounds) with 15 μg of acceleration. These are very small forces. A study was undertaken to examine if a force of about 3 pounds at a distance of approximately 100 feet, the distance of the object to be attracted to the Space Station, could be exerted using magnetic loop antennas, conventional electromagnets, or superconducting magnets. The temperature in the Space Station, due to background radiation, is about 100°K which is below the critical temperature of the recently discovered high temperature superconductors (HTSC). Provided there is sufficient shielding from the sun, the HTSC could be used in the Space Station environment and could use cryogen (liquid hydrogen and oxygen) already aboard the Space Station for cooling.

Another problem examined was the retrieval of a drifting crewmember using a life line technique. With this technique, the crewmember shoots a small iron ball or plate with a monofilament line attached to it, at an energized electromagnet in the Space Station using a spring action gun. The magnet attracts the iron material as it gets closer and holds it. The crewmember then uses the line to move to the Station. The same technique could be used to retrieve a drifting tool.

MAGNETIC FIELDS AND FORCES

Magnetic Field Due to A Dipole

The magnetic flux density due to a magnetic pole of strength 'm' at a distance r is given by[1]

$$B = \frac{\mu_0}{4\pi} \frac{m}{r^2} \qquad (1)$$

where

$$\mu_0 = 4\pi \times 10^{-7}$$

and has the units T.m/A or N/A^2 or W/A.m,

where T stands for Tesla or Weber/m^2 and N for Newton. The pole strength 'm' has units of ampere-meter (A.m.). The force on a magnetic pole of strength 'm' placed in the magnetic field is given by

$$F = m B \qquad (2)$$

The flux density at a distance much larger than the length of a dipole along its axis is given by [2]

$$B = \frac{\mu_o}{4\pi} \frac{2M}{r^3} \qquad (3)$$

where M is the magnetic moment of the dipole given by the product of its pole strength and length.

Force Between Two Dipoles

The potential energy of a dipole placed in the B field of another dipole is given by [3,4]

$$U = -\vec{M} \cdot \vec{B} \qquad (4)$$

The translational force acting on the dipole is,

$$F = -\nabla U = \vec{M} \cdot \nabla \vec{B} \qquad (5)$$

In a uniform B field, there is no net translational force on a dipole as the gradient of B is zero.

Using (3) in (5) we get

$$F \simeq -\frac{\mu_o}{4\pi} \frac{6 M_1 M_2}{r^4} \qquad (6)$$

The negative sign in (6) indicates force of attraction.

Field Due to A Current Loop on Its Axis
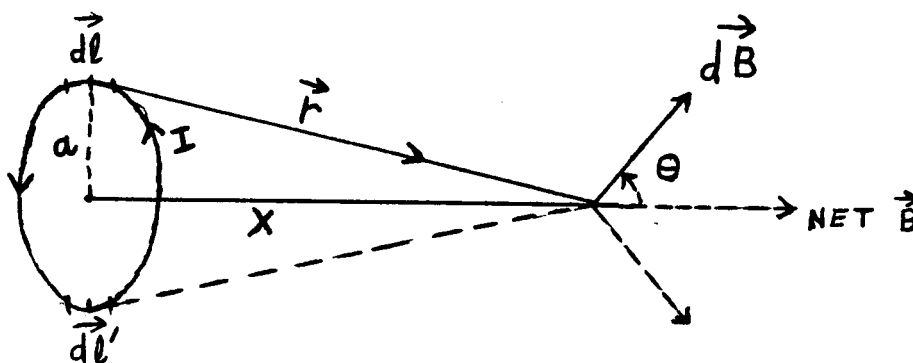
Figure 1.- Field On The Axis Of The Loop

The flux density at a point on the loop axis due to an element of length dl carrying current I is given by

$$d\vec{B} = \frac{\mu_0}{4\pi} \frac{I \, d\vec{l} \times \hat{r}}{r^2} \qquad (7)$$

where $\hat{r}$ is unit vector along $\vec{r}$. The field is normal to the vector $\vec{r}$ and element $d\vec{l}$.

The direction of field due to element $\vec{dl'}$ diametrically opposite to $d\vec{l}$ is shown. The components of the field at right angles to the axis cancel, while along the axis they add.

The net field along the axis due to a loop with N turns is

$$B = \int dB \cos\theta = \frac{\mu_0}{4\pi} \frac{IN\cos\theta}{r^2} \int_0^{2\pi a} dl$$

$$= \frac{\mu_0}{4\pi} \frac{2 I N A}{\left(a^2 + x^2\right)^{3/2}} \qquad (8)$$

for X >> a

$$B = \frac{\mu_0}{4\pi} \frac{2M}{X^3} \qquad (9)$$

where M = I N A $\qquad$ (10)

and has a direction given by right hand cork screw rule.

Comparison of (3) and (9) indicates that a current loop is equivalent to a magnetic dipole with its magnetic moment given by (10).

Force Between Two Current Loops

Let the magnetic moments of two loops be equal, that is, $M_1 = M_2 = I N A$.

If the two loops, each having a cross section area of $1m^2$, are separated by a distance r along their common axis, then the force between the two loops using (6) is given by

$$ F = \frac{\mu_0}{4\pi} \frac{6(IN)^2}{r^4} $$

which can be written in the form

$$ IN = \sqrt{\frac{4\pi F r^4}{6\mu_0}} \qquad (11) $$

A force of 15 Newtons (3.4 pounds) will move the OMV (weighing 20,000 pounds) with an acceleration of 150 μg and the Orbiter (weighing 225,000 pounds) with an acceleration of about 15 μg. The values of ampere-turns (IN) needed to provide a force of 15 Newtons between the two loops at various distances are calculated from (11) and the results are shown in Table 1. The values of ampere-turns needed at distances of 0.1m, 10m, and 30m are 50, 5×10^5, and 4.5×10^6, respectively, and they will generate a field of 80μT, 0.8 T, and 7.2 T, respectively, at the center of the loops.

TABLE 1.- DISTANCE VERSES AMP-TURN VALUES

r(m)	0.1	1.0	5.0	10	15	20	25	30
I N (Amp-Turn)	50	5×10^3	1.25×10^5	5×10^5	1.12×10^6	2×10^6	3.1×10^6	4.5×10^6

Torque on A Current Loop in A Uniform Field
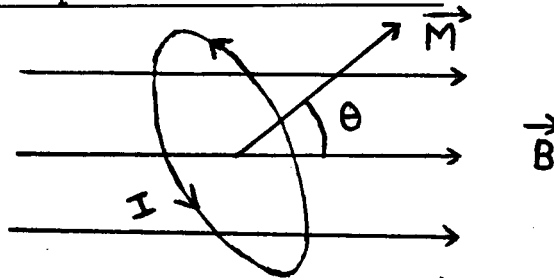
Figure 2.- Current Loop In A Uniform B Field

The torque acting on a current loop or a dipole placed in a uniform B field is given by [3,6]

$$\vec{T} = \vec{M} \times \vec{B} \qquad (12)$$

There will be a torque acting on the current loop due to the earth's magnetic field causing the loop to precess about the earth's field. The torque acting on the loop ($1m^2$) with 10^6 ampere-turn will have a maximum value of 35 N.m. If the loop is fixed in the spacecraft, then the spacecraft will precess about earth's field. This problem may be overcome by firing jets in the spacecraft that impart equal and opposite torque. Another solution to the problem of a single dipole in precessing in the earth's field is to use two dipoles with oppositely directed magnetic moment vectors (i.e., quadru-poles), such that the torques due to earth's field cancel out.

Field Due to A Quadrupole
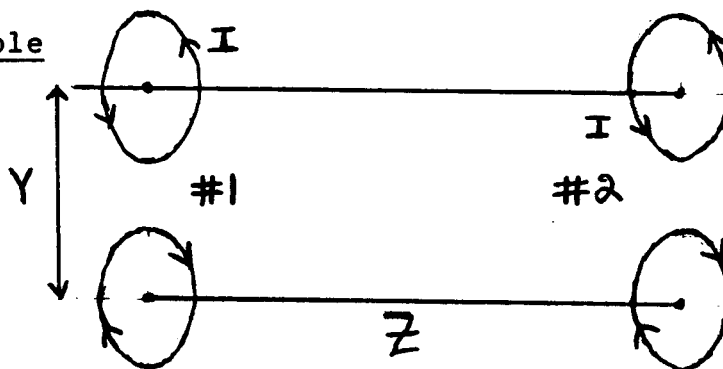
Figure 3.- Two Quadrupoles Separated By Z

The force that a quadrupole, consisting of a pair of current loops (with oppositely directed M vectors), separated by a distance Y, exerts on another quadrupole at a distance Z away is given by [5]

$$F \cong \frac{\mu_0}{4\pi} \frac{18 M_1 M_2 Y^2}{Z^6} \tag{13}$$

for Z >> y.

where M_1 is the magnetic moment of each of the dipoles in the quadrupole #1 and M_2 is corresponding value for quadrupole #2.

This configuration requires 4×10^7 ampere-turn in each of the four loops to generate a force of 15 Newtons at a separation of 30 meters between the two quadrupole and at a separation of 1 meter between the two loops of each of the quadrupoles. Thus, there is a high price to pay in going from dipole to a quadrupole system in terms of size and ampere-turns needed.

Force Between Current Loop and A Piece of Magnetic Material at A Distance

The force on a ferromagnetic material placed in the B field is given by

$$F = \vec{M} \cdot \vec{\nabla} B \tag{14}$$

where M is the induced magnetic moment of the magnetic material given by

$$M = \sigma V \tag{15}$$

σ is the induced magnetization in the magnetic material of Volume V.

In ferromagnetic material [1]

$$B = \mu_0 (H + \sigma) \tag{16}$$

also

$$B = \mu_r \mu_0 H \tag{17}$$

where μ_r is the relative permeability of the ferromagnetic alloy.

Since $\mu_r \gg 1$, then

$$\sigma \simeq B/\mu_0 \qquad (18)$$

The force of attraction between the current loop and the ferromagnetic alloy piece can be written using equations (14), (15), (18), and (9).

$$F = - \frac{3\mu_r \mu_0}{4\pi^2} \frac{M^2 V}{r^7} \qquad (19)$$

The force of attraction between a magnet and magnetic material decreases inversely as the seventh power of the distance.

$M = I N A =$ magnetic moment of the loop and $V = A_2 \delta X$, where δX is thickness of the iron plate.

For a current loop ($1m^2$) and a ferromagnetic alloy plate (4" x 4" x 1/8") with $\mu_r = 10,000$, the number of ampere-turns needed in the loop to produce an attractive force of one-tenth of a pound was calculated at various distances. The results obtained indicated values of 5.6×10^3, 1.56×10^6, and 1.77×10^7 ampere-turns at distances of 1m, 5m, and 10m, respectively.

When the magnet is in contact with the magnetic material, the force of attraction is given by [11]

$$F = B^2 A / 2 \mu_0 \quad \text{Newtons}$$

where A is the contact area and B is the flux density in the slight air-gap between the magnet and magnetic material. If B = 1.0 Weber/m^2, then F = 4×10^5 Newtons/m^2 or 58.0 pounds/in^2.

Weight of the Coil in the Current Loop

For ampere-turn value of 2×10^6 which will be needed to exert a force of 15 Newtons at 20m between the two current loops as well as between a current loop and sheet of ferromagnetic alloy at a distance of about 7m, let I be 10^4 amps and N = 200 turns. Assuming $J_c = 10,000$ Amp/cm^2 (projected value) in the high temperature superconductor (HTSC): Crossectional

area of the wire needed, $A = I/J_c = 1 \ cm^2$

$$= 10^{-4} \ m^2$$

length of the wire in the loop $= \pi d N$

$$= \pi (1m)(200) = 200 \pi \text{ meters}$$

Volume of the coil = area x length = $62.83 \times 10^{-3} \ m^3$

Weight of the coil = density x volumne = 440 Kgms = 970 pounds
The density value chosen is the average of cladding material
and the superconductor density ($7000 \ Kgm/m^3$). Length of the
loop = wire diameter x N

$$= \sqrt{\frac{4A}{\pi}} \times N = 11.3 \times 10^{-3} \times 200 =$$

2.26 meters.

The length of the loop could be reduced by increasing the
number of layers of the coil. To make a solenoid, the
superconductor needs to be copper stabilized in order to
conduct the heat in case of a quench (the superconductor
turning normal). The weight of the magnet will be 3-4 times
the weight of superconducting coil.

Radiation Field of A Magnetic Loop Antenna
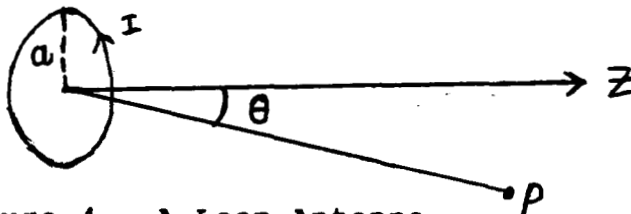
Figure 4.- A Loop Antenna

Let the normal to the plane of the loop of radius a be the
z-axis. The radiated H field due to small loop (a < $\lambda/6$) at a
point is given by [8,9]

$$H_\theta = \frac{\pi N A I \ SIN \ \theta}{\lambda^2 r} \qquad (20)$$

Using a phase locked loop in the incoming vehicle, the
current in the second magnetic loop in the vehicle is kept in
phase with the B field radiated by the first loop on the
Station. The force on the second loop due to B field of the
first loop antenna is

$$F = \vec{M_2} \cdot \vec{\nabla} B$$

If $\theta = 90°$, then

$$F = \frac{\mu_o \pi N_1 I_1 A_1 N_2 I_2 A_2}{r^a \lambda^a} \qquad (21)$$

22-10

assuming $A_1 = A_2 = 1m^2$, and $I_1 \, N_1 = I_2 N_2 = I \, N$, the equation (21) yields

$$I N = \frac{\lambda \, r}{2 \pi} \times 10^3 \sqrt{10 F} \qquad (22)$$

A force of 15 Newtons at a 30m separation between the two antennas will require 44×10^7 and 44×10^4 amp-turns at frequencies of 40 KHz and 40 MHz, respectively. The high value of amp-turns needed will require the use of superconducting loop. However, the eddy current losses at high frequency in superconductor will produce enough heat to render it normal.

The inductance of the loop coil is given by [10]

$$L = \mu_o \, N \, b \, log \left(b/a \right) \qquad (23)$$

where b and a are loop and wire radii, respectively. The reactance of the loop coil ($\omega \, L$) will be too high to achieve the required amp-turns with a moderate sized signal generator.

DISCUSSION

The recently discovered HTSC with critical temperature above 100°K have a demonstrated ability to remain superconducting in magnetic field of at least 30 Teslas. A review of high-field and high-temperature superconductors appeared in references 12-15. At present, the current density of bulk HTSC material is about 1000 amp/cm^2 but work is being done to increase it substantially by improving the material fabrication and processing techniques. The weight of superconducting coil in the current loop was calculated with a projected current density of 10,000 amp/cm^2; however, if the current density of the bulk HTSC material could be increased to 100,000 amps/cm^2, the weight of the superconducting coil would reduce to one-tenth of the calculated value. The current density of the thin film HTSC is about one million amp/cm^2. This high value makes one optimistic about reaching the goal for bulk material current density mentioned above.

In addition to the coil, a HTSC cylinder weighing about 20 pounds that can trap magnetic flux due to the Variant Incomplete Meisner Effect (VIME) can be used as a permanent magnet for fields up to 2.5 Teslas[5]. To make a permanent magnet cylinder using VIME, the HTSC cylinder is placed in the

magnetic field at $T > T_c$. Then the temperature of the cylinder is lowered below its T_c and the flux gets trapped in it, making it a permanent magnet as long as $T < T_c$. In case the HTSC cyclinder becomes normal and looses its magnetism, a magnetizing coil can be wound around to remagnetize it.

The use of conventional magnetic loop antennas with phase locked loop to provide a force of 15 Newtons at 30m distance was examined. It was concluded that due to inductance of the loop coil (0.2 mH for air core), the required value for ampere-turns would not be obtained with moderate sized signal generators.

To retrieve an adrift crewman, use of life line technique is proposed. The crewman would use a spring action gun that releases an iron ball or plate with a monofilament line attached to it towards a current loop that has been energized or towards the HTSC permanent magnetic cylinder in the Space Station. The iron ball or plate will be attracted to the loop or the magnet as it gets closer and will be held by it. Using the life line, the adrift crewmember will reach the Station. The 2.5 Tesla HTSC permanent magnet or loop with 2×10^6 amp-turns will be able to attract magnetic alloy plate (4" x 4" x 1/8") from a distance of 15-20 feet. The life line technique could be used to retrieve a tool that is drifting away. The crewmember will shoot a small permanent magnet at the tool with a line attached to it. The magnet will attach itself to the tool which can be pulled back using the line.

Jet propulsion will leave a lot of undesirable plumes hitting the Space Station when a spacecraft undocks from the Station. The ampere-turns needed to exert a repulsive force of about 3 pounds between the two current loops, one placed in the Space Station and the other in the spacecraft, will be 50, 5×10^3, and 1.25×10^5 at a distance of 0.1m, 1m, and 5m, respectively. The size of the current loops needed to provide the necessary force is not very large.

CONCLUSIONS

Current loops using the HTSC can be used to attract OMV or Orbiter with a force of 3 pounds at a distance of about 100 feet, although the size of the loop and the magnet is fairly large. With advances in material processing of HTSC, the current density could increase in the near future to a value that would reduce the size and weight of loop magnet substantially. A series of HTSC permanent magnets employing VIME could be used in place of loop magnets. Conventional magnetic loop antennas can not be used due to high reactance of

the loop at radiating frequencies. A life line technique could be used to rescue an adrift crewmember or tool in the early stages.

The size of the current loops needed for undocking a spacecraft from the Space Station is not very large. For undocking, magnetic repulsion would be used initially to move the spacecraft a few meters from the Station and then the jet propulsion resumed. This overcomes the problem of undesirable jet plumes hitting the Station.

Conventional magnets yielding field of several Teslas, such as a Bitter magnet cannot be used for space environment as the amount of power consumed by them is very large. The field of a Bitter magnet is given by [16]

$$B \simeq 11 \sqrt{W/r} \qquad \text{Teslas}$$

where W is the power consumed in megawatts and r is the radius of solenoid in cms. To produce a field of 11 Teslas with a conventional Bitter magnet with a bore size of 4 cms, will require consumption of two megawatts of power!

REFERENCES

1. Shortley, G., Williams, D., "Elements of Physics,"
    Prentice-Hall, Inc., Inglewood Cliffs, New Jersey (1965)

2. Vasudeva, D. N., "Fundamentals of Magnetism and
    Electricity," S. Chand & Co., New Delhi, India (1966)

3. Barnes, Thomas G., "Foundations of Electricity and
    Magnetism," D.C. Heath and Company, Boston, MA (1965)

4. Sears, Francis W., "Electricity and Magnetism,"
    Addison-Wesley Publishing Company, Inc., Reading, MA (1958)

5. Private Communications with Prof. Roy Weinstein of the
    University of Houston, University Park, Houston, TX

6. Scott, William T., "The Physics of Electricity and
    Magnetism," John Wiley & Sons, Inc., New York

7. Jackson, J. D., "Classical Electrodynamics," John Wiley and
    Sons, Inc., New York (1967)

8. Kraus, John D., "Antennas," McGraw-Hill Book Company,
    New York (1950)

9. Jasik, Henry, "Antenna Engineering Handbook,"
    McGraw-Hill, New York (1961)

10. Schelkunoff, Sergei, "Antenna Theory and Practice," John
    Wiley and Sons, Inc. New York (1952)

11. Hadfield, D., "Permanent Magnets and Magnetism," John Wiley
    and Sons, Inc., New York (1965)

12. Larbalestier, D., Fisk, Gene, Montgomery, B.,
    Hawksworth, D., "High-Field Superconductivity," Physics
    Today, March 86, pp 24-33

13. Fitzgerald, Karen, "Superconductivity: Facts vs. Fancy,"
    IEEE Spectrum, May 88, pp 30-41

14. Moon, Francis, Hull, John R., Berry, Greg F., "Super-
    conductivity: As Temperatures Rise, So Do Demand On MES,"
    Mechanical Engineering, June 88, pp 60-68

15. Rosen, Jerome, "Superconductivity: Obstacles Persist on
    the Path of No Resistance," Mechanical Engineering,
    June 88, pp 70-77

16. Parkinson, D. H., Mulhall, B. E., "The Generation of High
    Magnetic Fields," Plenum Press, New York (1967)

VISUAL PERCEPTION AND GRASPING FOR THE EXTRAVEHICULAR ACTIVITY ROBOT

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared By:	Scott A. Starks, Ph.D.
Academic Rank:	Professor
University & Department:	East Texas State University Computer Science Commerce, Texas 75428

ABSTRACT

Automation and robotics have played important roles in space activities for many years, most notably in planetary exploration and Space Shuttle operations. There are several major robotic systems being designed and developed for the Space Station. One such system is the Extravehicular Activity Retriever (EVAR). The EVAR will be an intelligent free-flying robot which will have the ability to locate, track, maneuver to, and retrieve objects which have been detached from the Space Station. Most notably, the EVAR is being designed so that it will be able to retrieve crew members who have become stranded during extravehicular activity. The EVAR is undergoing a series of ground-based demonstrations in the precision air bearing floor facility located at the NASA Johnson Space Center. These tests serve as a means of providing proof of concept as well as demonstrating system level performance of the EVAR hardware and software.

Space-based robotics systems are presented with many technical problems which can be avoided in robots designed for terrestrial settings. Due to the near absense of gravity in space, the manipulation of objects is quite different from what is normally experienced on earth. In contrast to the structured environments in which factory robots operate, space-based robots must deal with environments containing a high degree of uncertainty. The harsh nature of lighting in space creates visual perception problems as well. The end result is that although many robots have been developed which perform adequately on earth, much technology remains to be developed to support intelligent space-based robots such as the EVAR.

This study considers one such area of technology, the development of an approach to the visual perception of object surface information using laser range data in support of robotic grasping. This is a very important problem area in that a robot such as the EVAR must be able to formulate a grasping strategy on the basis of its knowledge of the surface structure of the object. This paper presents a description of the problem domain and formulates an algorithm which derives an object surface description adequate to support robotic grasping. The algorithm is based upon concepts of differential geometry namely, gaussian and mean curvature.

INTRODUCTION

It is the purpose of this paper to formulate an algorithm for the calculation of surface shape and structure from laser range data. This algorithm was developed with the intent of using it to assist in the closed loop control of the grasping operation performed by the Extravehicular Activity Retriever (EVAR). In order to perform the grasping operation a robot such as the EVAR must have knowledge of the structure of the surface of the object to be grasped. Using such information, the robot can plan the reach of its manipulator to the object and can preshape its hand in such a way so as to accommodate the object and ensure a stable grasp. The algorithm calculates information about the surface structure and is based upon concepts from differential geometry, namely the gaussian and mean curvatures of a surface. The approach to be presented is based in part upon previous works by other investigators [1], [2].

This report is based upon preliminary results from a study initiated this summer at the Johnson Space Center. In addition to the presentation of the algorithm for curvature-based surface calculation, this report also describes some of the key criteria relating to the determination of grasping strategies. It also presents a framework which allows a robot to store essential surface characterization and grasp configuration information in a computerized library. Under such a framework, a robot can make use of previously generated knowledge about objects and hand configurations to simplify the complexity of grasp planning.

EXTRAVEHICULAR ACTIVITY RETRIEVER

The use of robots in hazardous environments such as space is rapidly expanding. Numerous studies and symposia have been conducted on this very important topic [3]. The EVAR is a voice-supervised, intelligent, free-flying robot which is currently being designed, developed, and demonstrated in a ground-based laboratory facility at the Johnson Space Center [4]. When operational, the EVAR will be used to retrieve objects such as construction materials, equipment, and tools which have been accidentally separated from the Space Station. The EVAR will be in a stand-by mode whenever astronauts perform extravehicular activity (EVA). If an astronaut becomes detached from the Station, the EVAR will be summoned to locate, fly to, grapple and return the astronaut to the Station. The EVAR is being designed so that it can accommodate an unconscious or uncooperative astronaut.

The EVAR consists of six integrated subsystems: the Manned Maneuvering Unit (MMU), two robotic arms with grippers, a video and tracking system, a 3D laser imaging system, an onboard computer system, and a data and control network. The MMU provides the EVAR with the ability to fly about in a manner similar to that of an astronaut during EVA. The two robotic arms are anthropomorphic in nature. One of the grippers is a dexterous hand while the other is a parallel jaw gripper. The EVAR video and tracking system employs two black and white cameras, a tracking system, and a monitor to allow for backup supervisory control of EVAR operations by personnel at a remote location. One camera enables target acquisition and tracking, while the second can be used to assist during target grappling. The 3D laser imaging system is an Odetics 3D laser radar ranger. This subsystem, often called a 3D mapper, consists of a scan unit, an electronics unit, and a power supply. The 3D mapper provides direct digital range measurement and is capable of scanning a 60 degree horizontal by 60 degree vertical area with a 128X128 raster scan.

The frame rate of the system is every .835 seconds. The onboard computer system consists of several transputers and other processors. The data and control network enables the routing of communication and control information among the various subsystems of the EVAR. A three phase ground demonstration program for the EVAR has been planned and scheduled. The first phase of the program was successfully completed this past year. The second milestone demonstration will be conducted this fall at the precision air-bearing floor facility at NASA Johnson Space Center.

ROBOTIC GRASPING

Robot manipulation is a complex process that consists of several steps: sensing, task planning, trajectory planning, grasping, and path following. With respect to grasping, there are three principal grasp selection criteria which are of major concern. They are described briefly below.

Safety. The criteria of safety involves the issue of damage control as applied to three areas: (i) the target object must be safe in the sense that there should not be any damage to the object during the grasp operation; (ii) the robot must avoid damaging itself as it performs the operation; and (iii) the task environment should remain damage-free during the operation. Whereas the EVAR is being developed to retrieve stranded crew members and possibly very expensive Space Station construction tools and materials, requirement (i) is an overriding concern in the development of grasping strategies for the EVAR. In addition, the EVAR itself will be an expensive piece of equipment, therefore elevating the importance of requirement (ii).

Reachability. This criteria centers on the ability of the hand to reach a particular target. In other words, there must be a sufficiently large workspace to accommodate the movement of the hand during the operation. This requires that the hand be able to reach the object at some initial hand configuration and, with the object in hand, to follow a collision-free path to a final location. In addition, the workspace of the hand's fingers must be large enough relative to the size of the object for the fingers to secure the proper grasp.

Stability. The grasp must be stable in the presence of external forces which could be exerted on the grasped object of the hand during the manipulation process. By stable, we simply mean that a grasp does not permit the target to move or slip with respect to the hand.

It is clear that the above grasp criteria are closely related to the characteristics of the targets and the environments which they occupy. In fact, they form the bases upon which grasp configurations and operation strategies are determined. In the following section, the characteristics of targets which are most critical to the grasping operation are presented.

CHARACTERISTICS OF OBJECTS

Objects can be characterized by three different types of attributes (geometric, physical, and mechanical) [5]. Geometric attributes include the size of an object, the shape of the object, and the shape of the contact surface. With regard to grasping, the size of the object is most important in determining the accessibility of the object. The shape of the object as well as that of its contact surface influence grasp stability and hand preshaping. Prior to selecting a grasp, a list of candidate grasp configurations can be extracted from a library. Candidate grasp configurations can then be evaluated on the basis of the geometric attributes of the object. The set of candidate grasp configurations could then be narrowed by removing those grasp configurations that would lead to unstable grasps. Finally, the physical and mechanical characteristics of the object could be taken into consideration.

Physical attributes include the mass, mass distribution, and inertia of the objects. The

center of mass affects the stability of the grasp configurations. The physical attributes of the target are difficult, if not impossible, to extract using solely non-contact methods of sensing, such as imaging or laser ranging. This being the case, it would be advantageous to develop a library containing the physical properties of objects likely to be retrieved by the EVAR [6]. By classifying the target, the EVAR could then retrieve its physical attributes and make use of these in the formulation of a grasp strategy.

Target objects can be grouped into five categories based upon mechanical attributes. These categories are rigid, brittle, elastic, flexibile, and slippery objects. Rigid objects exhibit a high degree of stiffness which makes them unlikely to deform during grasping. Unlike rigid objects, brittle objects have low stiffness. Potentially, objects of this type can be very fragile and thus easily damaged. Grasping brittle objects requires accurate information concerning the shape and size of contact areas in that they can be easily damaged. Elastic objects are characterized on the basis of behavior of their surfaces about the point of contact by a manipulator. When a force is exerted on an elastic object, deformation of the surface at the point of contact of the manipulator may appear. When the force is removed, the deformation disappears, restoring the object to its original form. Elastic objects are typically the easiest objects to grasp. Unlike elastic objects, flexible objects have low stiffness. As a result, deformation which result from the application of a force remain subsequent to the removal of the force. Normally, detailed shape information about the target is not as important for grasping elastic objects. Large contact areas are essential for grasping flexible objects. Slippery objects are perhaps the most difficult to grasp and handle. Whereas the coefficient of friction at the contact area is very low, the stability of the grasp can be diminished.

It is important to note that the mechanical attributes of objects should be considered in conjunction with geometric and physical attributes in the determination of grasp strategies. Such design ensures that any drawback in one aspect can be compensated by advantages offered by the others.

VISUAL PERCEPTION OF RANGE IMAGING

In a broad sense, visual perception of range images can be viewed as the process of interpreting measurements made using any of a variety of range sensors. Because the success of automated devices depends critically on systems with the ability to sense and understand the environment, range image perception has received the attention of many investigators. Automatic systems have been designed to work in tightly structured environments such as assembly lines in factories. Unfortunately, much of the technology associated with systems of this type is not directly transferrable to the EVAR in that the EVAR must have the ability to cope with uncertainities in its environment.

The EVAR is equipped with a laser ranging device which is capable of determining a range image of the environment surrounding the robot. A range image is merely a large collection of distance measurements from a known reference coordinate system to surface points on objects within a scene. If the distance measurements in a range image are listed relative to three orthogonal coordinate axes, the range image is said to be in xyz form. If the distance measurements indicate range along 3D direction vectors indexed by two integers (i, j), the range image is said to be in r_{ij} form. Any image in r_{ij} form can be converted directly to xyz form, but the converse is not true [7]. The term image is used because any r_{ij} range image can be displayed on a video monitor, and it is identical in form to a digitized video image from a television camera. In addition the term visual perception is often associated with the extraction of information from a range image for the same reason.

Range images can be a source of valuable information for an automated machine. The EVAR can use range image information to determine the location and orientation of objects in its environment. This information can be used to plan the EVAR's movements and actions. Surface fitting techniques make use of range data to produce a geometric description of objects. In this report, we are interested in determining the underlying surface structure of targets which the EVAR will be responsible for grasping. An approach to extracting the surface structure of a target is critical to the grasp operation because manipulators grasp surfaces. The following section presents an approach to surface fitting which is based upon curvature concepts taken from the field of differential geometry.

SURFACE FITTING

The surface fitting problem can be stated as follows: given three m x n arrays of x, y, and z coordinates, we would like to determine a surface that approximates the data in the least-squares sense and that is smooth. In a mathematical sense, smoothness is defined as the twice-differentiability of the surface at all points. Fitting a surface to an L x L window of data is equivalent to computing a surface fit for a roughly rectangular grid of data values. Intuitively, a roughly rectangular grid is one that has been obtained from a rectangular mesh that has been deformed to fit the surface. We may represent the surface in parametric form by the three equations

$$x = f(s,t),$$

$$y = g(s,t),$$

$$z = h(s,t),$$

where s and t are the parameters, and the functions f, g, and h are tensor products of splines in tension. The fitting can be viewed as a mapping from 2D space to 3D space. The 2D space is characterized by the parameters s and t and the 3D space is the standard cartesian space.

Principal curvatures at a point on a surface indicate how fast the surface is pulling away from its tangent plane at that point. Principal curvatures can be computed by estimating partial derivative infomation for the surface. The principal curvatures will achieve a local maximum in the area surrounding a discontinuity in the surface, for example an edge in the object. By thresholding the values of the principal curvatures, we can declare the edge points in a range image.

When a plane passing through the normal to the surface at a point P is rotated about this normal, the radius of curvature changes and will be a maximum distance r_1 for a definite normal section s_1 and a minimum r_2 for another normal section s_2. The reciprocals $k_1 = 1/r_1$ and $k_2 = 1/r_2$ are called the principal curvatures; the directions of the tangents to s_1 and s_2 at P are called the principal directions of the surface at P. Gaussian curvature at point P is defined as the product of the two principal curvatures. It can be proven that the gaussian curvature depends only upon the coefficients of the first fundamental form of a surface and their derivatives. The coefficients of the second fundamental form share the previously stated property. The first and second fundamental coefficients also determine the surface uniquely up to a rigid body transformation. As a

result, the gaussian curvature is said to be an intrinsic property of the surface. The principal curvatures at a point on a surface can be computed in terms of the parameters s and t using the following approach. Let X(s,t) represent the surface

$$X(s,t) = [\ x(s,t),\ y(s,t),\ z(s,t)\ ]^T.$$

Here, bold face notation is used to denote a vector quantity. If we let partial differentiation be represented by subscripts, the differential element dX is a vector given by the relationship

$$dX = X_s\ ds + X_t\ dt,$$

where

$$X_s = dX/ds = [\ x_s,\ y_s,\ z_s\ ]^T,$$

$$X_t = dX/dt = [\ x_t,\ y_t,\ z_t\ ]^T.$$

If we take the scalar product of dX with itself we obtain a relationship which is known as the First Fundamental Form of the surface, I, which is given below:

$$I = dX \cdot dX = (\ X_s\ ds + X_t\ dt\ ) \cdot (\ X_s\ ds + X_t\ dt\ )$$

$$= E\ ds^2 + 2F\ dsdt + Gdt^2,$$

where the First Fundamental Coefficients (E, F, and G) are given by:

$$E = |\ X_s\ |^2 = x_s^2 + y_s^2 + z_s^2,$$

$$F = X_s \cdot X_t = x_s\ x_t + y_s\ y_t + z_s\ z_t,$$

$$G = |\ X_t\ |^2 = x_t^2 + y_t^2 + z_t^2.$$

The unit normal is then given by

$$N = (\ X_s\ x\ X_t\ )\ /\ |\ X_s\ x\ X_t\ |.$$

The differential of the unit normal is

$$dN = N_s\ ds + N_t\ dt.$$

If we take the negative of the scalar product of the differentials of the surface and the unit normal, we obtain the Second Fundamental Form for the Surface,

$$II = -dX \cdot dN = -\ (X_s\ ds + X_t\ dt) \cdot (N_s\ ds + N_t\ dt) =$$

$$= L\ ds^2 + 2\ M\ ds\ dt + N\ dt^2,$$

where the Second Fundamental Coefficients (L, M, and N) are given by:

$$L = -X_s \bullet N_s,$$

$$M = -1/2 \ (X_s . N_t + X_t . N_s),$$

$$N = - X_t \bullet N_t.$$

The gaussian curvature K at any point on the surface is defined as the product of the two principal curvatures k_1 and k_2 and can be expressed in terms of the First and Second Fundamental Coefficients as

$$K = k_1 k_2 = (LN - M^2) / (EG - F^2).$$

The mean curvature H is given by the average of the two principal curvatures. It may be expressed in terms of the First and Second Fundamental Coefficients as well,

$$H = (LG - 2MF + NE) / (2(EG-F^2)).$$

The principal curvatures are given by,

$$k_1 = H - SQRT(H^2 - K),$$

$$k_2 = H + SQRT(H^2 - K).$$

The above formulas can be used to compute the gaussian and mean curvatures. Let us assume that the gaussian and mean curvatures for a particular point referenced by the indices (i, j) in a range image have been calculated and are stored in arrays $K(i, j)$ and $H(i, j)$ respectively. The points on a surface can then be classified according to the signs of these quantities: positive, negative, or zero. It turns out that there are only eight possible outcomes for surface characterization based upon the signs of the gaussian and mean curvatures. We introduce notation for two functions (modified signum function and the surface classification function). The modified signum function is a mapping of a scalar argument into one of three values (1 , 0 , or -1). For positive e, it is defined as:

$$sgn_e \ (y) = \begin{array}{ll} 1 & if \ \ y > e \\ 0 & if \ \ |y| < e \\ -1 & if \ \ y < -e. \end{array}$$

We take advantage of the clarity afforded by use of the modified signum function to introduce the surface classification function

$$T(i,j) = 1 + 3(1 + \text{sgn}_{e1}(H(i,j))) + (1 - \text{sgn}_{e2}(K(i,j))).$$

The positive scalars $e1$ and $e2$ are chosen to be close to zero. They enable the mapping to zero of a number within the limits of the numerical accuracy of the machine and the algorithm. Figure 1a gives the mapping of the values of the modified signum function applied to the gaussian and mean curvatures. Figure 1b gives a pictorial interpretation of the eight possible surface classifications afforded by the function $T(i,j)$.

CURVATURE-BASED SURFACE CALCULATION ALGORITHM

This section presents an algorithm to compute the object description in terms of jump boundaries, internal edges, and regions homogeneous in the sign of their gaussian and mean curvatures. The algorithm can be summarize as a sequence of the following steps:

• Divide the range image into overlapping windows.
• Detect jump boundaries and fit patches to windows of data not containing jump boundaries.
• Compute the principal curvatures and extract edge points.
• Classify each nonedge point in a patch as one of the eight possible surface classifications presented previously.
• Group all points of the same type in a patch and its neighboring patches into a region.

The details of each step in the algorithm are as follows. In the first step, the division of the input arrays of (x, y, z) coordinates into L x L windows should be overlapped to ensure that an internal edge is always contained in a patch.

In the second step, the relative degree of scatter exhibited by the data is reflected in the standard deviation of the euclidean distance between adjacent data points within the L x L window. In the vicinity of a jump boundary of an object, the standard deviation of these euclidean distances will be relatively high. Jump boundaries can be detected by means of a threshold test. After the jump boundaries have been detected, smooth patches based upon 2D B-splines can be fitted to the data in the window not containing a jump discontinuity [8]. In the presence of a jump boundary, the window size can be modified to ensure that the window excludes the jump boundary. In order to detect fine detail in the scene it is important that the size of the objects be much larger than the size of the L x L window.

In the third step of the algorithm, we attempt to detect all those points that belong to or fall on internal edges of the object. As was mentioned previously, in the vicinity of an edge, the values for the principal curvatures will be high and in fact achieve a local maximum. As a result, we first determine all points exhibiting principal curvatures above threshold. Due to the presence of noise in the data and an inappropriate choice of thresholds, clusters of edge points may appear in the vicinity of a true edge position. To eliminate these clusters of computed edge points, a suppression of nonmaxima is applied at every edge point. That is, we only declare an edge to be present at points where the curvature is a local maximum. Nonmaxima suppression is applied in a direction perpendicular to the edge directions which is the direction associatated with the maximum absolute principal curvature.

The next step is to group object points into homogeneous regions. All points in a

T(i,j)	Surface Type
1	Peak
2	Ridge
3	Saddle Ridge
4	Not possible
5	Flat
6	Minimal
7	Pit
8	Valley
9	Saddle Valley

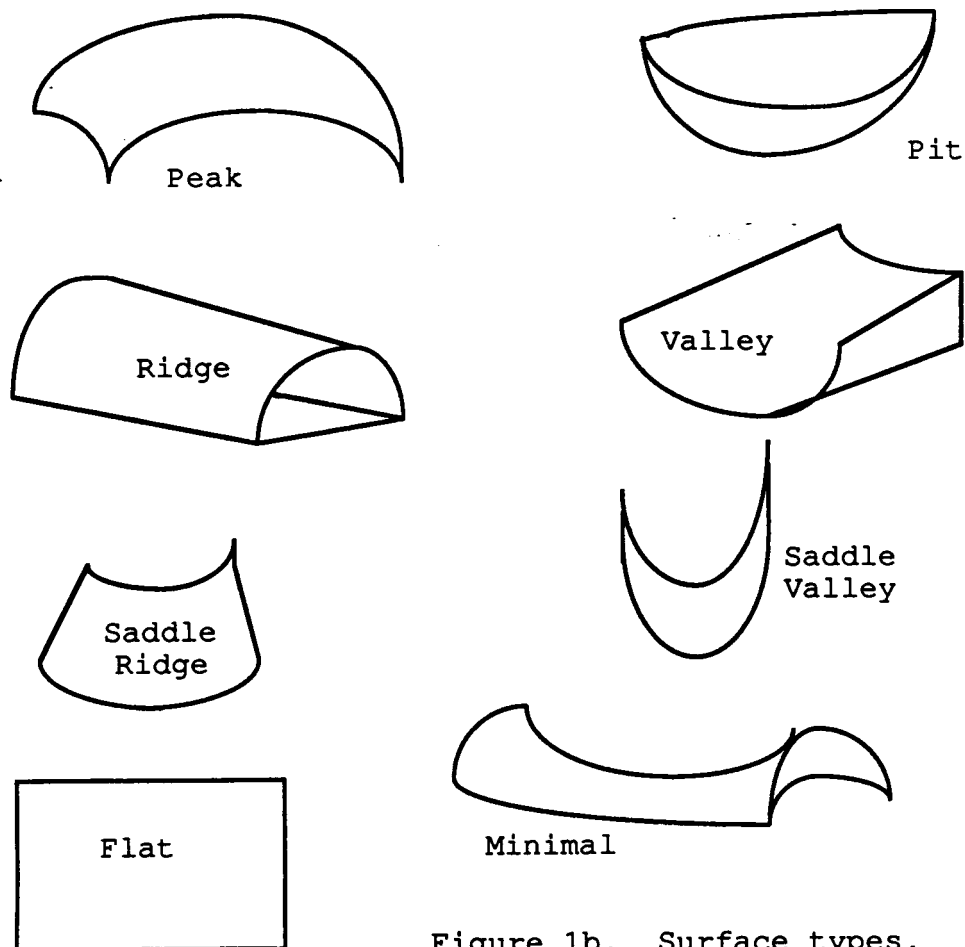Figure 1a.   Classification of surfaces

Figure 1b.   Surface types.

patch are classified into one of the types indicated in Figure 1b. All points of the same type are grouped together into a larger region. Ther merging of points does not require explicit fitting of a new surface to this homogeneous region since the surfaces are chosen to be smooth. Each homogeneous region is assigned a label depicting its type. Internal edges may occur within regions. The extent of regions can be delineated by jump boundaries, internal edges, and curvature edges, which we define as places where there is a change in curvature-based classification.

The object representation in terms of regions and curvature-based properties , as described above, have the following advantages:
• it is invariant under transformation of independent parameters of the surface.
• it is invariant to rigid body transformation and is therefore independent of viewpoint.

EVAR GRASPING

Operational constraints of the EVAR Demonstration Program demand that grasping be performed in a real-time, closed-loop manner. Under Phase II of the program, both video and range data will be available for the demonstration. In addition, the target will be a stationary object. It is assumed that the EVAR hand will be in close proximity (within a foot) of the target prior to grasping.

Probably the single most important property of the human hand is its ability to grasp a wide range of objects. The incorporation of a hand on the EVAR will enable the EVAR to grasp a wide variety of objects as well. Grasping involves the coordination of many degrees of freedom. In an automated system, we wish to reduce the number of degrees of freedom associated with robotic grasping. One approach to achieve this goal is based upon the notion of classification of grasp configurations. By dealing with grasp configurations in a packaged format through classification, one can simplify the amount of planning associated with the grasping operation [9]. This approach is based upon the abilty of the automated system to recognize surface structures based upon features derived from image and range measurements. Features could be constructed using the curvature-based surface representation calculation algorithm to describe the underlying suface structure of the object to be grasped. These features could be matched against those stored in a surface/grasp library [10]. This library would contain grasp configuration which would relate to the particular surface structure. In addition, the library would contain the physical and mechanical attributes of the associated object. Suppose the EVAR is presented with a surface structure similar to one previously encountered and successfully grasped. Under such circumstances, the EVAR can make use of information previously generated concerning grasping to sieze the target. This information would consist of hand preshape geometry, grasp approach, and other relevant information.

If the surface cannot be classified by matching its surface features with those contained in the surface/grasp library, it will be necessary to formulate a new grasp strategy. In that the hand must grasp surfaces, the information generated using the algorithm would be useful in the development of a new grasp configuration. If a new configuration is developed and successfully used to produce a stable grasp, then it is stored along with the underlying target surface structure in the surface/grasp library. Needless to say, the surface/grasp library will be dynamic. As new objects are successfully grasped, the library will expand. The overall approach is illustrated in Figure 2.

SUMMARY

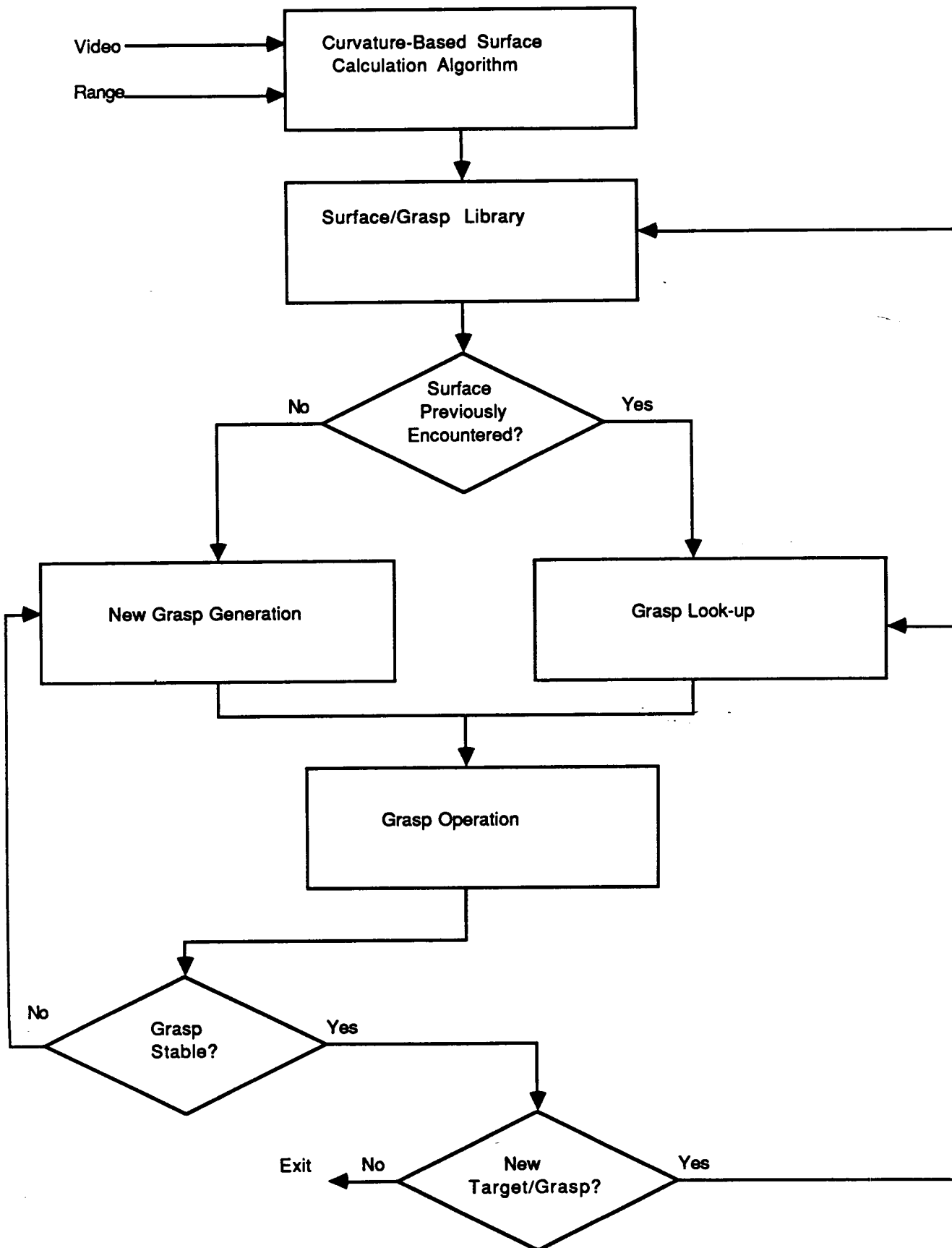This paper has presented an algorithm for the calculation of curvature-based surface

Figure 2. Framework for grasp operation.

23-12

characteristics for objects using range data. At the present time, this algorithm is being implemented in the programming language C. Although the approach to surface description based upon curvature concepts taken from differential geometry appears promising, it would be presumptuous to leave the reader with the impression that it will provide all the information necessary for secure and stable grasping for the EVAR. It is essential to integrate this prototype software, when available, in a real-time systems level demonstration in order to fully evaluate its merits. Such a demonstration would bring to light any deficiencies which could then be addressed by future refinements.

REFERENCES

1. B.C. Vemuri, A. Mitiche, and J.K. Aggarwal, "Curvature-based representation of objects from range data," *Image and Vision Comput.*, vol. 4, no. 2, pp. 107-114, May 1986.

2. P.J. Besl and R.C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 10, no. 2, March 1988.

3. J. Erickson, "Manned spacecraft automation and robotics," *Proc. of the IEEE*, vol. 75, no. 3, March 1987.

4. G.J. Reuter, C.W. Hess, D.E. Rhoades, L.W. McFadin, K.J. Healey, and J.D. Erickson, "An intelligent, free-flying robot," to appear in *Proc. of SPIE: Space Station Automation IV*, November 1988.

5. G. Wang and H.E. Stephanou, "Chopstick manipulation with an articulated hand: a qualitative analysis," *Proc. of 1988 IEEE Intl. Conf. on Robotics and Automation*, pp. 94-102, 1988.

6. W.C. Chiou and S.A. Starks, "An introduction to the concept of robot factors and its application to space station automation," *Proc. of SPIE: Space Station Automation*, vol. 580, pp. 53-57, September 1985.

7. P. J. Besl, "Active, optical range imaging sensors," *Machine Vision and Appl.*, vol. 1, no. 2, pp. 127-152, 1988.

8. W. Tiller, "Rational B-splines for curve and surface representation," *IEEE Comp. Graphics and Appl.*, pp. 61-69, September 1983.

9. D.M. Lyons, "A simple set of grasps for a dextrous hand," *Proc. of 1985 IEEE Intl. Conf. on Robotics and Automation*, pp. 588-593, 1988.

10. K. Rao, G. Medioni, H. Lui, and G.A. Bekey, "Robot hand-eye coordination: shape description and grasping," *Proc. of 1988 IEEE Intl. Conf. on Robotics and Automation*, pp. 407-411, 1988.

ESTIMATION OF VELOCITY PERTURBATIONS IN SATELLITE FRAGMENTATION EVENTS

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1988

Johnson Space Center

Prepared by: Arjun Tan

Academic Rank: Associate Professor

University & Department: Alabama A&M University
                         Department of Physics
                         Normal, Alabama 35762

ABSTRACT

The magnitude, variance and directionality of the velocity perturbations of the fragments of a satellite can shed valuable information regarding the nature and intensity of the fragmentation. Up until now, the only method used to calculate the three orthogonal components of the velocity change consisted of inverting the process of evaluating the changes in the orbital elements of the fragments due to velocity perturbing forces. But the traditional method failed in five different cases: one, when the parent satellite's orbit was circular; two and three, when the true anomaly of the parent was either 0° or 180°; and four and five, when the argument of latitude of the parent was 90° or 270°. This report describes a new method of calculating the velocity perturbations which is free from the shortcomings of the traditional method and could be used in all occasions, provided the fragmentation data and the orbital elements data are consistent with one another.

The method uses the parent satellite's local frame of reference at the time of breakup. The three orthogonal components of the velocity change are derived from three simultaneous equations provided by the changes in specific energy, specific angular momentum and plane angle of the fragment. The fragmentation parameters and the orbital elements of the fragmented satellites are taken from the satellite fragmentation catalog while the orbital elements of the fragments are read from the NORAD and PARCS data files. In order to reverse the decay of the fragment's orbit, the orbital elements of the fragment are propagated backwards to the time of fragmentation.

The velocity perturbations of the fragments of over 20 major satellite fragmentation events were calculated using the new method in this study. This method, together with a recent technique of determining the masses of the fragments, have now furnished a complete set of information about the fragments, including the mass, velocity, momentum, effective cross-sectional area and diameter. The magnitudes and variances of the velcotity perturbations of the fragments of Himawari rocket and Cosmos 839 satellite have reaffirmed their classifications as low intensity explosion and collision induced breakups respectively. The preferred directionality of the fragments of Landsat 3 rocket indicates that the rocket most likely broke up in the "clam" model of low intensity explosion. The residual velocity changes of the fragments of Solwind P-78 satellite provide indications of the general direction of the incoming interceptor. The differential velocity distribution of the fragments of Delta class of breakups could be fitted with a beta function, whereas the cumulative distribution of the same followed a general power law. There is generally a negative correlation between the mass and the velocity change of the fragments. A similar correlation exists between the effective diameter and the velocity change.

INTRODUCTION

To date, nearly one hundred artificial Earth satellites have fragmented in orbit, thus contributing a large proportion of unwanted hazardous material in space (Johnson and Nauer, 1987). To determine retrospectively the cause of fragmentation from the debris characteristics is one of the principal goals of the debris researchers. Over the past few years, we have witnessed significant progress in this area. From the debris data, Culp and McKnight, 1985, developed a satellite fragmentation test to distinguish between explosion and collision induced fragmentation events. Based on the distribution of plane change angle and radar cross-sections of known fragmentation events, Badhwar, et al., 1988, devised a scheme to distinguish between the three classes of satellite breakups: low intensity explosion induced, high intensity explosion induced and hypervelocity collision induced breakups. Furthermore, a technique now exists to determine the masses of the fragments of a satellite from the history of their orbital elements and radar cross-sections (Badhwar and Anz-Meador, 1988).

One of the most important quantities in satellite fragmentation events is the velocity change suffered by a fragment during the breakup. For instance, the magnitudes of these velocity changes may be used as indications as to whether a fragmentation is due to low intensity explosion, high intensity explosion or hypervelocity impact, based on the study of Bess, 1975. Also, the directionality of the velocity perturbations may give clues as to the nature and intensity of the explosion as per the analysis of Benz, et al., 1987. However, up until now, no satisfactory method existed to calculate all three components of the velocity change. The traditional method consisted of inverting the process of evaluating the changes of the orbital elements due to velocity perturbing forces from Meirovitch, 1970, and had been used by McKnight, 1986; Kling, 1986; and Tan, 1987. But Tan, 1987, showed that the calculated radial component of the velocity change became exceedingly large when the parent satellite's orbit was nearly circular or when the fragmentation took place near the apsidal points. Similarly, the calculated cross-range component of the velocity change became singularly high when the argument of latitude was 90° or 270°.

In order to obtain a method which is free from the singularities of the traditional method and which could be used in all occasions, Reynolds, 1987, initiated a new approach of estimating the velocity changes. This approach parallels an earlier study by Wiesel, 1978, and was refined by Badhwar, NASA scientist; Tan, summer fellow; and Reynolds, Lockheed scientist at NASA-JSC this summer. The method is now fully operational and coupled with the technique of Badhwar and Anz-Meador, 1988, can furnish a complete set of information about the fragments, including the mass, velocity, momentum, effective cross-sectional area and diameter. A brief description of the method is contained in this report. The report also summarizes the results obtained by analyzing the velocity perturbations in the major breakup events calculated by this method.

METHOD

In order to calculate the velocity perturbations imparted to the fragments, it is convenient to use the parent satellite's local frame of reference at the point of breakup (Wiesel, 1978; Reynolds, 1987). The three orthogonal directions are then defined by the radial and the down-range directions in the plane of the orbit and the cross-range direction perpendicular to the plane of the orbit and along the angular momentum vector (Fig.1). In this coordinate system, the velocity $\vec{v}$ of the parent has the components v_r, v_d and 0. From elementary astrodynamics, we have

$$v = [\mu(\frac{2}{r} - \frac{1}{a})]^{\frac{1}{2}} , \tag{1}$$

$$v_d = \frac{1}{r} [\mu a(1-e^2)]^{\frac{1}{2}} , \tag{2}$$

and $\quad v_r = \pm (v^2-v_d^2)^{\frac{1}{2}} , \tag{3}$

where a is the semi major axis and e the eccentricity of the parent's prbit, r is the radial distance to the breakup point and μ is the gravitational parameter. In Eq.(3), the plus sign corresponds to the ascending mode of the satellite (true anomaly $\nu < 180°$) whereas the minus sign corresponds to descending mode ($\nu > 180°$).

Upon fragmentation, the velocity of a fragment $\vec{v}'$ will have the components v_r+dv_r, v_d+dv_d and dv_x, where dv_r, dv_d and dv_x are the velocity perturbations received by the fragment during the breakup. In consequence, the fragment would suffer changes in energy, angular momentum and orbital inclination. The change in specific energy is given by

$$E'-E = \frac{1}{2}(v'^2-v^2) = - \frac{\mu}{2} (\frac{1}{a'} - \frac{1}{a}) ,$$

or, $\quad \frac{1}{2}[(v_r+dv_r)^2 + (v_d+dv_d)^2 + dv_x^2 - v_r^2 - v_d^2] = - \frac{\mu}{2} (\frac{1}{a'} - \frac{1}{a}) , \tag{4}$

where a' is the semi major axis of the fragment's orbit. Similarly, if e' is the eccentricity of the fragment's orbit, the change in specific angular momentum is given by

$$h'^2-h^2 = |\vec{r} \times \vec{v}'|^2 - |\vec{r} \times \vec{v}|^2 = \mu a'(1-e'^2) - \mu a(1-e^2) ,$$

or, $\quad r^2[(v_d+dv_d)^2 + dv_x^2 - v_d^2] = \mu a'(1-e'^2) - \mu a(1-e^2) . \tag{5}$

Finally, the plane change angle θ of the perturbed orbit from the unperturbed is given by (vide Fig.2)

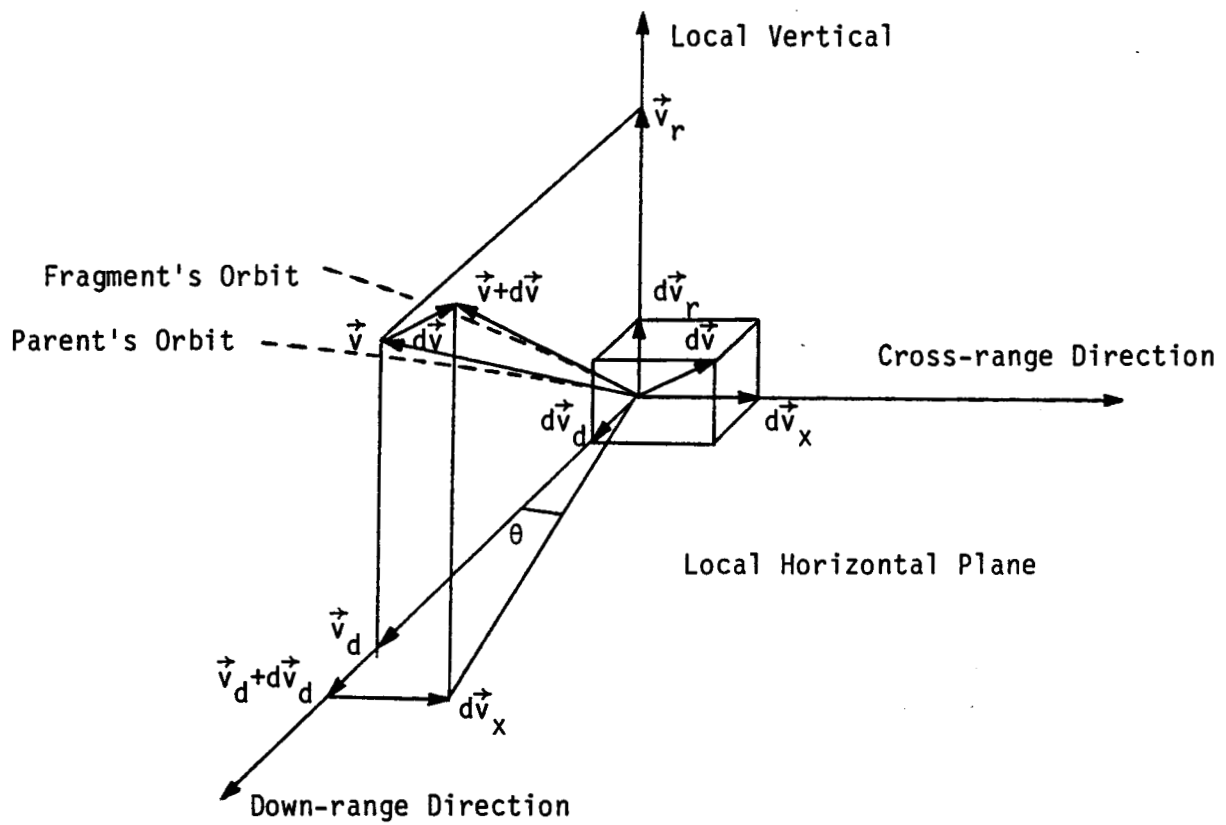$$\tan \theta = \frac{dv_x}{v_d+dv_d} . \tag{6}$$

Fig.1. The parent satellite's local coordinate system and the velocity components of the parent and the fragment.
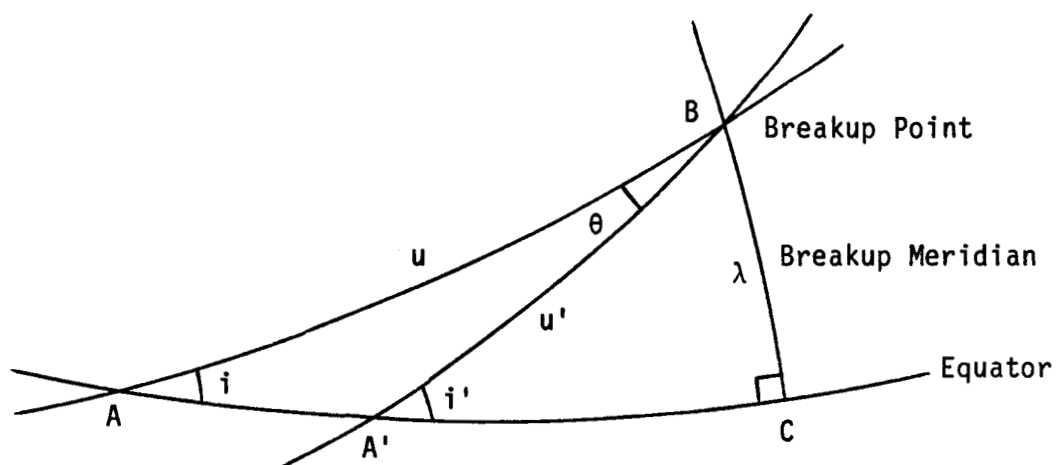

Fig.2. Breakup (spherical) triangles for the parent and the fragment.

Equations (4), (5) and (6) comprise three simultaneous equations from which the three unknowns dv_r, dv_d and dv_x can be solved in principle. After some manipulations and elimination, we arrive at the expressions for the velocity perturbations

$$dv_r = \pm \left[\mu\left(\frac{2}{r} - \frac{1}{a'}\right) - \frac{\mu}{r^2} a'(1-e'^2)\right]^{\frac{1}{2}} - v_r \ , \tag{7}$$

$$dv_d = \frac{\cos\theta}{r} \left[\mu a'(1-e'^2)\right]^{\frac{1}{2}} - v_d \ , \tag{8}$$

and $\quad dv_x = \frac{\sin\theta}{r} \left[\mu a'(1-e'^2)\right]^{\frac{1}{2}} \ . \tag{9}$

In Eq.(7), the plus sign corresponds to the ascending mode of the fragment (true anomaly $v' < 180°$), whereas the minus sign corresponds to descending mode ($v' > 180°$). Elsewhere, the dominance of v_d over dv_d ensures the retention of only the positive sign of the square root in Eqs.(8) and (9). Equations (7), (8) and (9) can be shown to be equivalent to Reynolds' (1987) equations, which are expressed in normalized quantities.

The plane change angle θ can be expressed as a function of the inclinations i and i' of the parent's and the fragment's orbits respectively and λ, the latitude of the breakup point (cf. Badhwar, et al., 1988). Applying both the cosine laws to the spherical triangle AA'B of Fig.2 and the sine law to the triangles ABC and A'BC and substituting, we get

$$\theta = \pm \cos^{-1} \frac{\cos i \cos i' + (\cos^2\lambda - \cos^2 i)^{\frac{1}{2}}(\cos^2\lambda - \cos^2 i')^{\frac{1}{2}}}{\cos^2\lambda} \ . \tag{10}$$

Here, the plus sign corresponds to i' > i and the minus sign corresponds to i' < i.

The true anomaly v' of the fragment at the time of breakup, which dictates the sign of $v_r + dv_r$ in Eq.(7) is determined from the argument of latitude u' and the argument of perigee ω' of the fragment at the time of breakup as follows.

$$v' = u' - \omega' \ . \tag{11}$$

Applying the sine law to the spherical triangle A'BC of Fig.2, we get

$$u' = \sin^{-1}\left(\frac{\sin\lambda}{\sin i'}\right) \quad \text{or} \quad 180° - \sin^{-1}\left(\frac{\sin\lambda}{\sin i'}\right) \ , \tag{12}$$

where the first solution corresponds to northbound motion of the fragment and the second corresponds to southbound motion at the time of fragmentation. Since the argument of perigee is perturbed by the oblateness of the earth, the argument of perigee of the fragment at the time of observation ω'_o is different from that at the time of fragmentation ω'. From the rate of precession derived by King-Hele, 1964, we get

$$\omega' = \omega'_0 - \frac{4.98(5 \cos^2 i' - 1)\Delta t}{(a'/r_0)^{7/2}(1-e'^2)^2} ,$$

(13)

where Δt is the time of observation from the time of fragmentation. Here, ω', ω'_0 are expressed in degrees and Δt is expressed in days. It was verified from the successive NORAD data that the true anomaly predicted this way was generally accurate to within 2 degrees per year of the observed value, thus demonstrating the validity of this technique.

The fragmentation parameters and the orbital elements of the fragmented satellites are taken from the fragmentation catalog by Johnson and Nauer, 1987, while the orbital elements of the fragments are read from the NORAD data files. The apogee and perigee heights are normally converted into the semi major axis and eccentricity. Thus

$$a = \frac{2r_0 + h_a + h_p}{2} ,$$

(14)

$$e = \frac{h_a - h_p}{2r_0 + h_a + h_p} ,$$

(15)

$$r = r_0 + h ,$$

(16)

$$a' = \frac{2r_0 + h'_a + h'_p}{2} ,$$

(17)

and $$e' = \frac{h'_a - h'_p}{2r_0 + h'_a + h'_p} ,$$

(18)

where r_0 is the reference radius of the earth, h_a and h_p are the apogee and perigee heights of the parent, and h'_a and h'_p are those of the fragment.

Atmospheric drag causes decay in satellite orbits, particularly at low altitudes, making both the semi major axis and eccentricity smaller. This introduces substantial errors in the velocity perturbations calculations when the debris data were measured long after the fragmentation event, which is quite often the case. In order to compensate for this error, the debris parameters are propagated backwards to the time of fragmentation by running the orbital decay code of Mueller, 1981 in reverse. For this purpose, the area-to-mass ratio of the fragments are first determined by the technique of Badhwar and Anz-Meador, 1988. The solar flux in that period, which determines the atmospheric densities, is read from the NOAA data base.

RESULTS

In this study, over 20 major satellite fragmentation events were studied, each of which produced at least 40 cataloged objects. The fragmentation parameters and the orbital elements of the parent satellite were taken from the satellite fragmentation catalog by Johnson and Nauer, 1987 and the orbital elements of the fragments were read from the NORAD data files. In addition, the PARCS data for the Solwind P-78 fragments were used.

Figure 3 shows the frequency distributions of the velocity perturbations of the fragments of Himawari rocket body and Cosmos 839 satellite. The frgmentation of the Himawari rocket belongs to the Delta class of fragmentation due to low intensity explosion, whereas Cosmos 839 was believed to have been an ASAT target which was destroyed through collision (cf. Johnson and Nauer, 1987). The magnitudes of the velocity changes are much smaller in the latter case as compared with the former, which lend support to the classifications based on the study of Bess, 1975. Further, the velocity changes in the Cosmos 839 fragmentation event have a much smaller variance compared with those of the Himawari explosion, which are again consistent with the classifications (vide Badhwar and Tan, 1988).

Figure 4 depicts the two- and three-dimensional plots of the velocity perturbations of the fragments of Solwind P-78 satellite which was intentionally destroyed through hypervelocity impact (Kling, 1986). The velocity perturbations show considerable scatter and their magnitudes are significantly larger than those in the alledged Soviet ASAT collisions. Conspicous in the figures are three fragments with large velocity changes in all three directions.

In order to study the directionality of the fragments, we define two angles as follows: the colatitude $\theta = \cos^{-1}(dv_r/dv)$, $0° < \theta < 180°$; and the longitude $\phi = \tan^{-1}(dv_x/dv_d) + n\,180°$, $-180° < \phi < 180°$, where $n = 0$ for $dv_d > 0$, $n = 1$ for $dv_d < 0$, $dv_x > 0$, and $n = -1$ for $dv_d < 0$, $dv_x < 0$. Note that the colatitude is not the plane change angle defined earlier. The angles for each fragment of a breakup event are calculated and then plotted. Fig.5a is an angular plot of the fragments of Landsat 3 rocket as in a Mercator projection of the world map. The preferred direction in which most fragments emerged clearly shows that the distribution of fragments was highly anisotropic in this case. Based on the analysis of Benz, et al., 1987, we infer that Landsat 3 was most likely broken in the "clam" model of low intensity explosion. Fig.5b is a similar plot with the velocity changes in addition. Since the Mercator plots are area-distorted with the polar fragments appearing less dense compared with the equatorial, we have calculated the fragment flux per unit solid angle, where the elementary solid angle is $d\Omega = \sin\theta\,d\theta\,d\phi$. Fig.5c exhibits the fragment flux per unit solid angle, which is now free from area distortion.
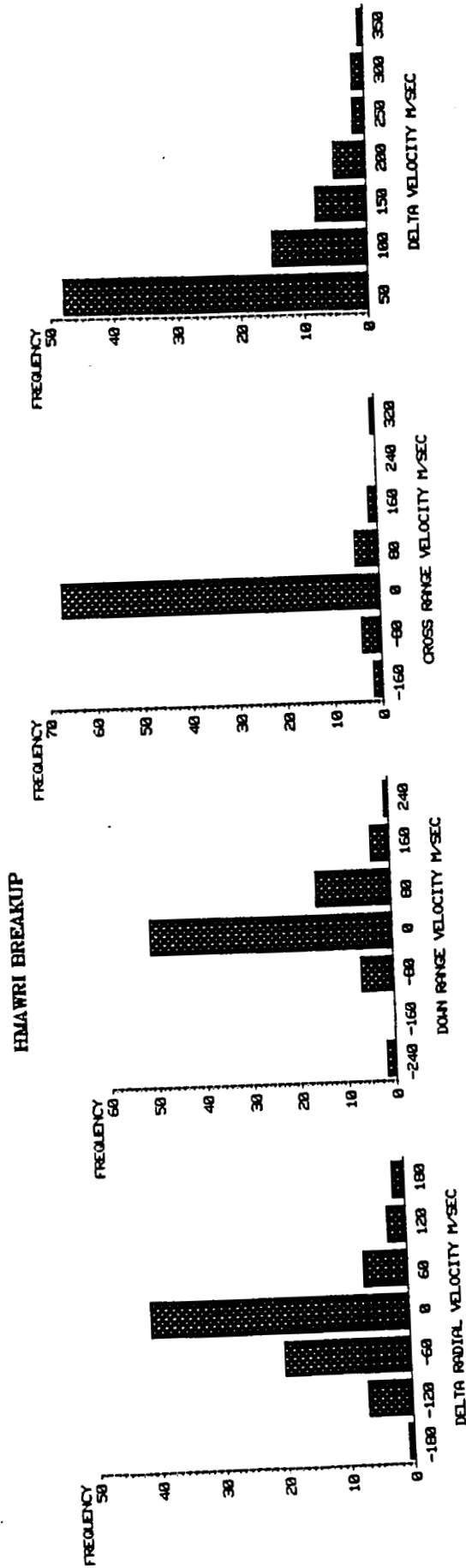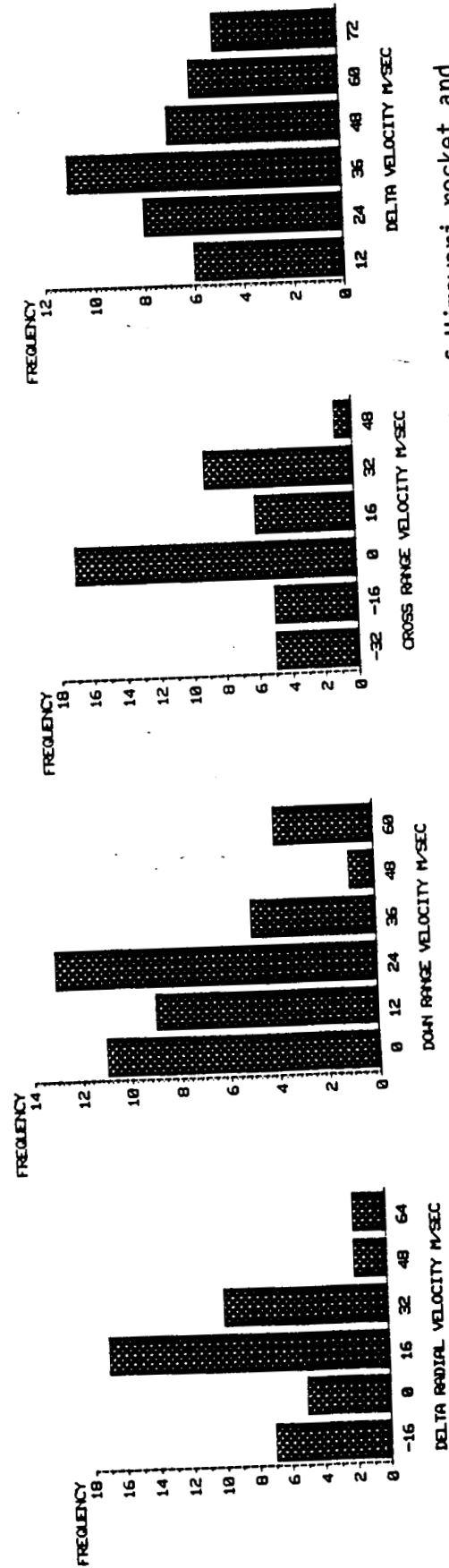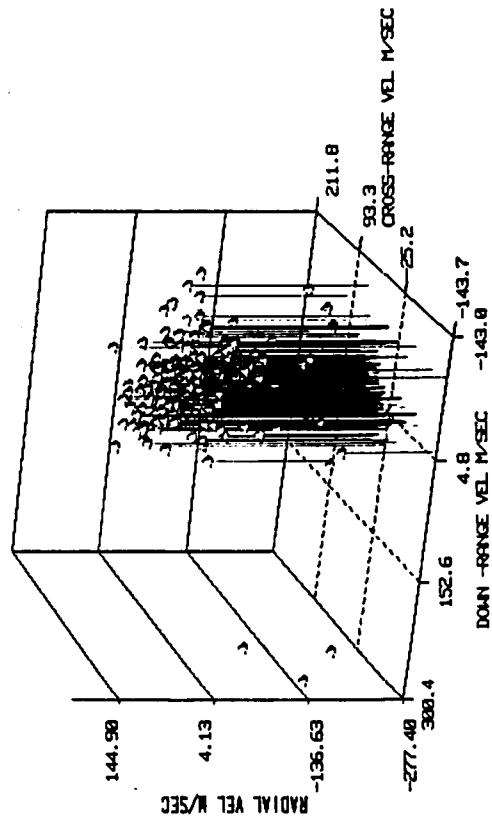
Fig.3. Frequency distributions of the velocity perturbations of the fragments of Himawari rocket and Cosmos 839 satellite.
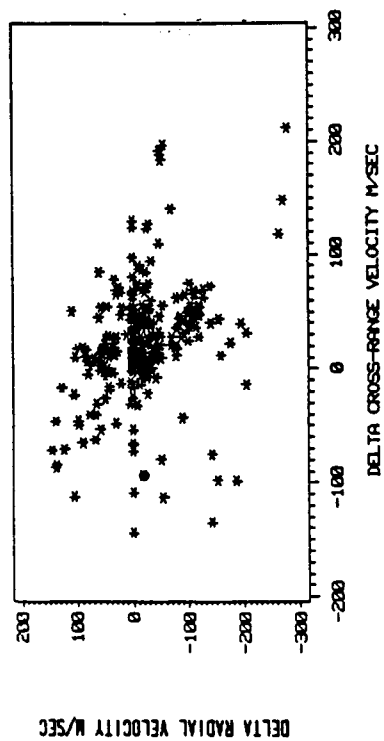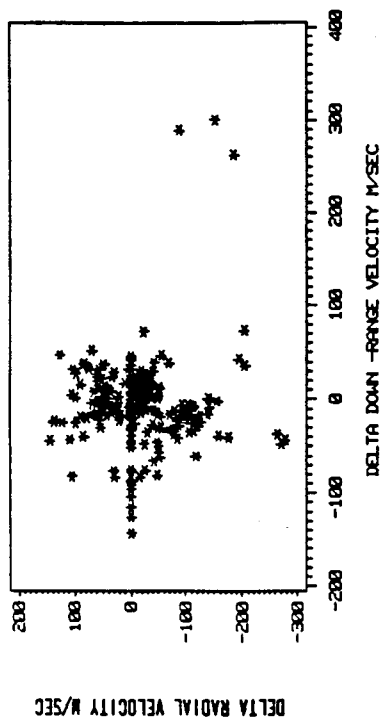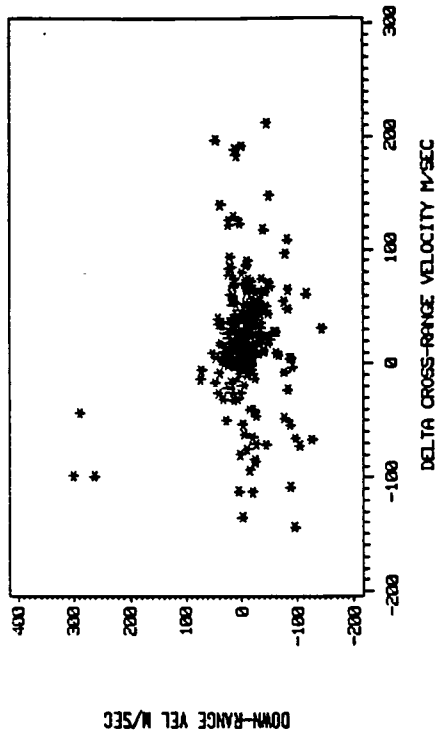
NEWP78 BREAKUP

Fig.4. Two- and three-dimensional plots of the velocity perturbations of the fragments of Solwind P-78 satellite.
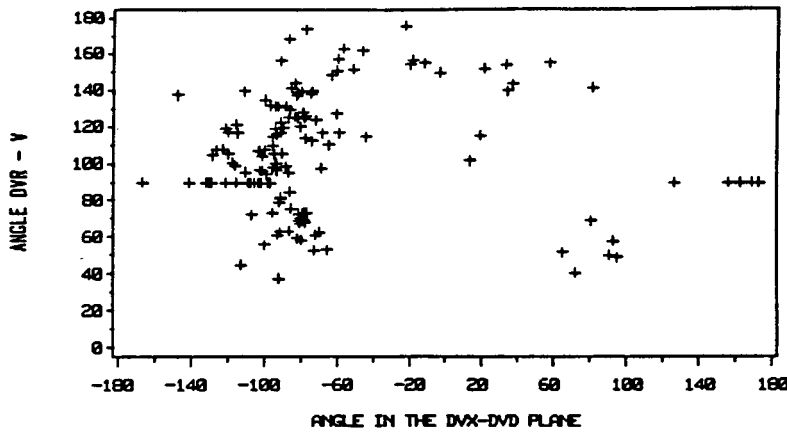
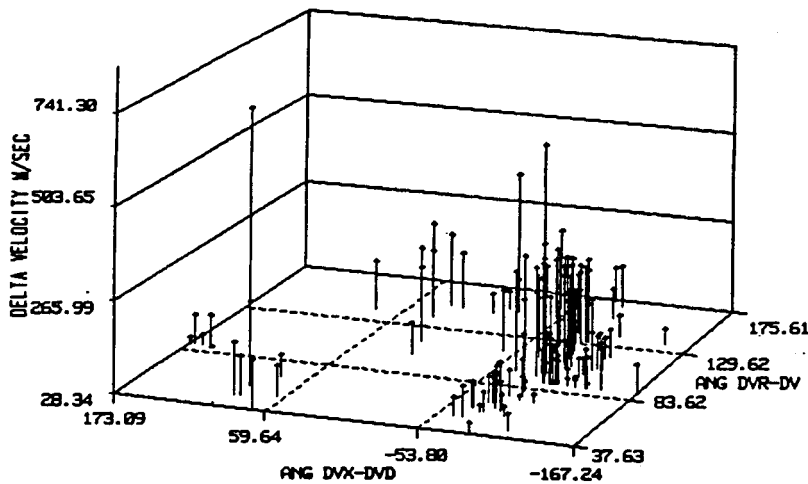LNDST3 BREAKUP

Fig.5a.
Angular
distribution
of fragments
of Landsat 3
rocket.

Fig.5b.
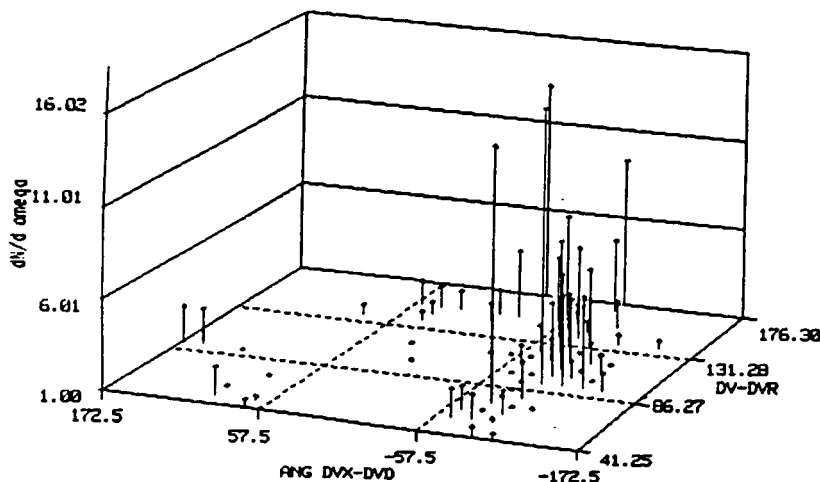Angular plot
of fragments
of Landsat 3
with velocity
change.

Fig.5c.
Distribution
of fragment
flux per unit
solid angle.

Analysis of the Solwind P-78 fragmentation event with the PARCS data taken immediately after the fragmentation shows that the mean velocity changes of all fragments, i.e., the mean dv_r, dv_d and dv_x do not add up to zero but were -17.6, -4.9 and 22.5 meter per second respectively. For the lack of orbital element history, the masses of the fragments could not be determined. However, assuming equal masses for all fragments, we find a net transfer of momentum in the direction of $\theta = 103°.1$ and $\phi = 102°.3$, which could imply that the interceptor most probably arrived from the general direction of $\theta = 76°.9$ and $\phi = -77°.7$ in the local frame of reference of Solwind P-78.

Figure 6a shows the differential distribution of the fragments of all Delta rocket breakups (NOAA, Landsat and Himawari) versus velocity change. The curve, which shows considerable extension in velocity, can be fitted with a beta function of the form

$$\frac{dN}{d(dv)} = A \, (dv/dv_{max})^{\alpha-1} \, (1 - dv/dv_{max})^{\beta-1} \, . \tag{19}$$

Fig.6b is a plot of the cumulative number of fragments of the above category of breakups with velocity change greater than dv versus the velocity change. The cumulative distribution gives a smoother curve which is given by a power law of the form

$$N(>dv) = A \, (dv + B)^{-C} \, . \tag{20}$$

The constants α, β, A, B, C etc. which determine the shapes of the distributions, could be used as characteristic parameters of an individual or a class of breakup events.

The correlation, if any, between the masses and the velocity changes of the fragments could be found once the masses of the fragments are determined by the technique of Badhwar and Anz-Meador, 1988. Fig.7 is a plot of mass against the velocity change of the fragments of the Spot-1 Viking rocket. Amidst considerable scatter, the trend for a negative correlation is clearly noticeable.

The change in momentum suffered by a fragment during the breakup can be computed from the masses and the velocity changes thus determined. Fig.8 is an example of the cumulative distribution of the fragments of NOAA 5 rocket versus momentum change. The data points were fitted with an expression similar to Eq.(20).

The effective diameter of the Delta class of fragmentations have been computed from the effective cross-sectional area and plotted against the velocity change with error bars in Fig.9. Finally, the distribution of the fragments of the same class of events are plotted against the velocity change and effective diameter in Fig.10. A pyramid-shaped distribution with unequal slopes on the opposite sides is clearly visible. This distribution can be used in the study of the evolution of debris cloud after a model low intensity explosion.
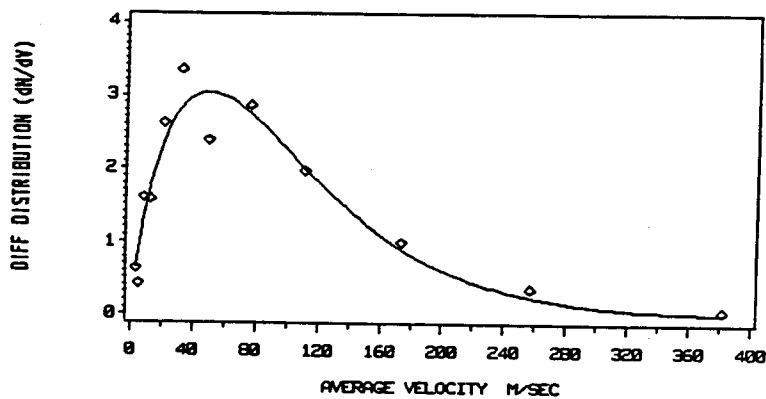
DELTA CLASS VELOCITY DISTRIBUTION

Fig.6a.
Differential
distribution
of fragments
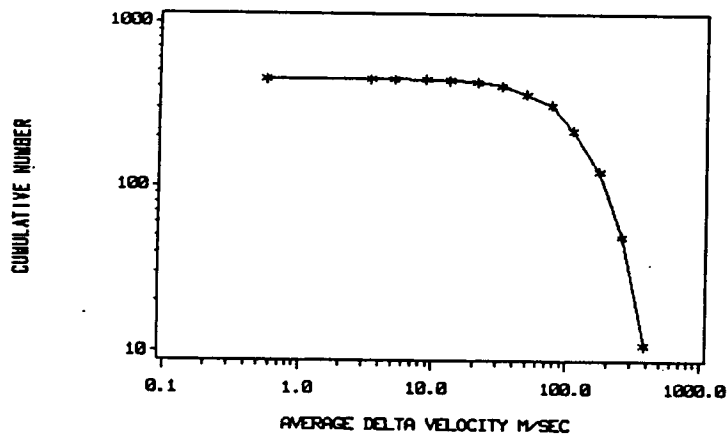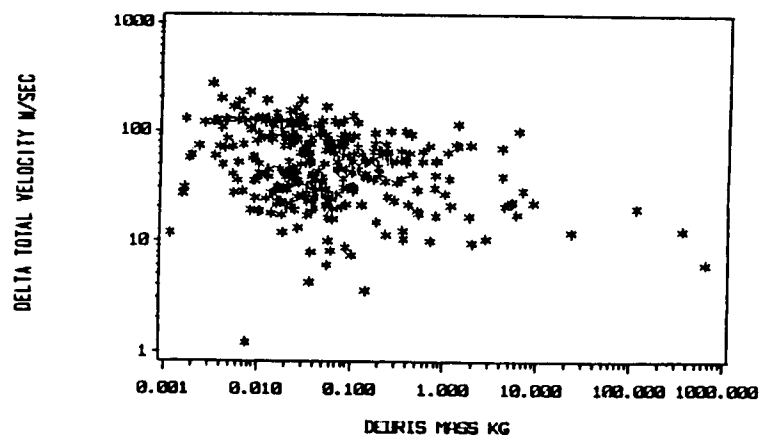of Delta class
fragmentations
vs. velocity
change.

Fig.6b.
Cumulative
distribution
of fragments
of Delta class
fragmentations
vs. velocity
change.

SPTVIK BREAKUP

Fig.7.
Velocity
change of
fragments
of Spot-1
Viking rocket
vs. fragment
mass.

24-13

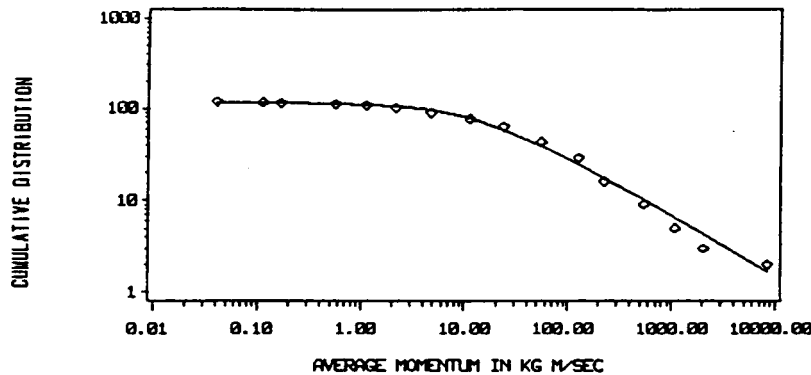## NOAA5 MOMENTUM DISTRIBUTION

Fig.8.
Cumulative
distribution
of fragments
of NOAA 5
rocket vs.
average
momentum
change.

DELTA CLASS BREAKUP

Fig.9.
Velocity
perturbations
of fragments
of Delta class
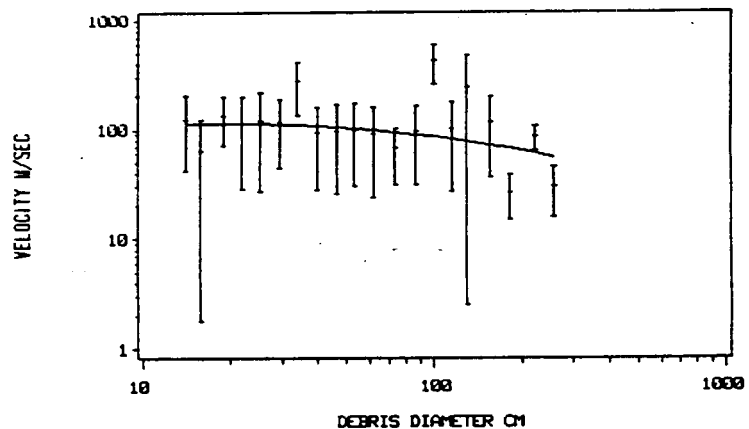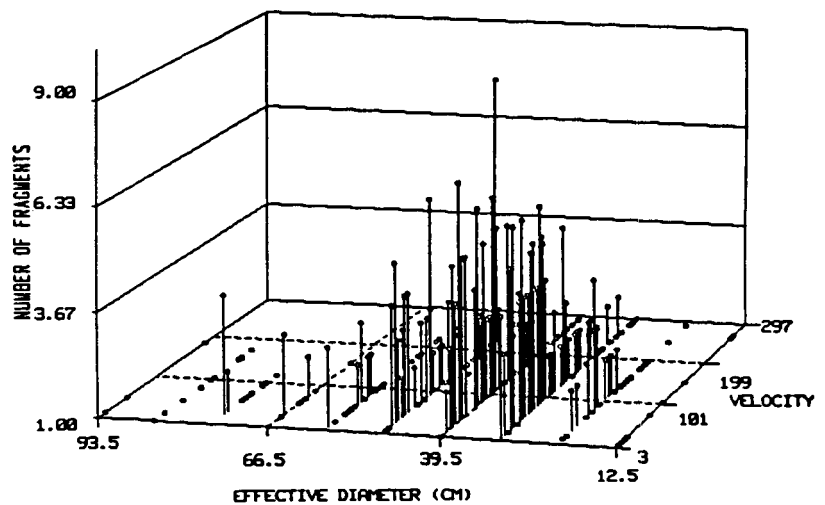fragmentations
vs. effective
fragment
diameter.

Fig.10.
Distribution
of fragments
of Delta class
fragmentations
with velocity
change and
effective
fragment
diameter.

24-14

REFERENCES

Badhwar, G. D.; and Anz-Meador, P. D.: Determination of the area and mass distribution of orbital debris fragments, preprint, 1988.

Badhwar, G. D.; and Tan, A.: Analysis and interpretation of satellite fragmentation data, J. Ala. Acad. Sci., Vol. 59, 1988, p. 153.

Badhwar, G. D.; Potter, A. E.; Anz-Meador, P. D.; and Reynolds, R. C.: Characteristics of satellite breakups from radar cross-section and plane change angle, J. Spacecraft and Rockets, Vol. 44, July 1988.

Benz, F. J.; Kays, R. L.; Bishop, C. V.; and Eck, M. B.: Explosive fragmentation of orbiting propellant tanks, unpublished, 1987.

Bess, T. D.: Mass distribution of orbiting man-made space debris. NASA TN D-8108, 1975.

Culp, R. D.; and McKnight, D. S.: Distinguishing between collision-induced and explosion-induced satellite breakup through debris analysis. Paper presented at AAS/AIAA Astrodynamics Specialist conference, Vail, Colorado, 1985.

Johnson, N. L.; and Nauer, D. J.: History of on-orbit satellite fragmentations, Teledyne Brown Engg. Report CS88-LKD-001, 1987.

King-Hele, D.: Theory of satellite orbits in an atmosphere, Butterworths, 1964, p. 4.

Kling, R.: Postmortem of a hypervelocity impact, Teledyne Brown Engg. Report CS86-LKD-001, 1986.

McKnight. D. S.: Discerning the cause of satellite breakups. Paper presented at the 62nd Annual AAS meeting, Boulder, Colorado, 1986.

Meirovitch, L.: Methods of Analytical Dynamics, McGraw-Hill, 1970, pp. 453-456.

Mueller, A. C.: The decay of low earth satellite, Lockheed Engg. Report 17520, 1981.

Reynolds, R. C.: Evolution of the debris cloud immediately after a satellite breakup in orbit, unpublished, 1987.

Tan, A.: Analysis and interpretation of satellite fragmentation data in NASA CR 172009, W. B. Jones and S. H. Goldstein ed., 1987.

Wiesel, W.: Fragmentation of asteroids and artificial satellites in orbit, Icarus, Vol. 34, 1978, pp. 99-116.

TIME-DOMAIN IMAGING

Final Report

NASA/ASEE Summer Faculty Fellowship Program 1988

Johnson Space Center

Prepared By: C. L. Tolliver, Ph.D., P.E.

Academic Rank: Professor

University & Department: Prairie View A&M University
                                            Department of Electrical Engineering
                                            Prairie View, Texas 77447

NASA/JSC

Directorate: Engineering

Division: Tracking and Communications

Branch: Systems Analysis Office

JSC Colleague: Kumar Krishen, Ph.D

Date: August 19, 1988

Contract Number: NGT 44-005-803

Table of Contents

Abstract

The quest for the highest resolution microwave imaging and principle of time-domain imaging has been the primary motivation for recent developments in time-domain techniques. With the present technology fast time varying signals can now be measured and recorded both in magnitude and in-phase. It has also enhanced our ability to extract relevant details concerning the scattering object. In the past, the interface of object geometry or shape for scattered signals has received substantial attention in radar technology. Various scattering theories were proposed to develop analytical solutions to this problem. Furthermore, the random inversion, frequency swept holography, and the synthetic radar imaging, have two things in common: (a) the physical optic far- field approximation and (b) the utilization of channels as an extra physical dimension, were also advanced. Despite the inherent vectorial nature of electromagnetic waves these scalar treatments have brought forth some promising results in practice with notable examples in subsurface and structure sounding. The development of time-domain techniques are studied through the theoretical aspects as well as experimental verification. The use of time-domain imaging for space robotic vision applications has been suggested.

I. Introduction

The time-domain imaging is the synthesis of the scattered electromagnetic field distribution over an object plane. Over the past few years, an interest has developed in the use of time- domain methods or data in order to solve scattering problems. This approach has many attractive features, among them the large amount of information contained in a single pulse response and simple physical interpretations; and improvement in time-domain technology increases the interest in such an approach. The purpose of this paper is to present a short review of some techniques that have been used in the analysis of time-domain scattering. The problem of interest is the reconstruction of the shape of a convex scatter from knowledge of high frequency far field scattered from the object in response to a known incident field.

The quest for the highest resolution microwave imaging and the principle of time-domain imaging have been the primary motivation for recent developments in time-domain techniques. In the last decade, modern techniques in sampling devices and the advent of fast pulse generators have brought new technology to the practicing radar engineers in the measurements of picoseconds at a greatly reduced cost. With the present new technology, fast time varying signals can now be measured and recorded both in magnitude and in-phase.

The goals of NASA are to study the reflection coefficient at and below the surface of the tiles on the Space Shuttle with special emphasis on time-domain robotic vision applications and to continue research into problems related to the exploration of space and its applications of new technology and techniques. To obtain these goals, the Space Station shall be designed as a multi-purpose facility in which missions of long duration can be conducted and supported. These missions will include science and applications, observation, technology development and demonstration, commercial laboratories and production facilities, operational activities such as servicing/maintenance, repair of satellites, support of unmanned platforms, assembly of large space systems, and as a transportation node for transfer to other orbits and planetary missions. Automation and Robotic (A&R) is rapidly growing and it is an important technical area foreseen to increase productivity and enhance astronaut safety. The use of A&R for the Space Station can be viewed in two major areas: (1) teleoperated/robotic systems for servicing maintenance, repairs, and assembly, and (2) computerized systems to reduce the manpower requirements of planning, monitoring, diagnosis, control, and fault recovery of systems/subsystems.

In addition to increased productivity through autonomy, the A&R will result in increased operational capability and flexibility. Robotic operations for the Space Station will involve maintenance/repair of the entire structure including various subsystems, orbiter/satellite servicing, astronaut assistance, equipment transfer, docking and berthing, inspection, remote monitoring, rocket staging, telescience, and assembly of the station and large structures. To aid the astronaut in various tasks and replace him/her for some activities, robots must perform beyond the current state-of-the-art by responding to a high degree of environmental uncertainty and operational flexibility. In order to accommodate various performance goals in robots, design concepts have been proposed by Dr. Kumar Krishen of NASA and other researchers.

Today's technology enables us to extract pertinent details concerning the scattering object which conventional radar ranging lacks. The inference of object geometry or shape from scattered signals has received substantial attention in radar

technology largely due to its academic significant and the understandable commercial and military values. In the past decade, various scattering theories were proposed to develop analytical solutions to this problem. It is well known that random inversion, frequency swept holography, and the synthetic radar imaging have two things in common: (a) physical optic far-field approximation, and (b) utilization of the channels as an extra physical dimension.

Despite the inherent vectorial nature of electromagnetic waves, these scalar treatments have brought forth some promising results in practice with notable examples in subsurface and structure sounding. Recently, Dr. Kumar Krishen, of NASA/JSC, has proposed the use of time-domain imaging for space robotic vision applications. A multi-sensor approach to vision has been shown to have several advantages over the video approach.

II. Theory

The theoretical approach outlines the basic principles of time-domain imaging relative to the synthesis of the scattered electromagnetic field distribution over an object plane. It was stressed that it is the scattered field, not the total field that is of interest, it is known by virtue of the boundary conditions, the tangential electrical field must be zero at a perfect conducting surface, where as that of the magnetic field is discontinuous by the surface current. The literature review reveals that the theory will not change and it is applicable to any arbitrary source waveform.

The scattering problems from rough surfaces is inherently different in nature from that of scattering by other bodies. Usually in the rough surface problem, an exact knowledge of the shape of the surface is neither available, nor is it of interest to the radar operator. Instead, only average properties of the surface shape enter into the problem. It is clearly understood that the last requirement rules out a boundary-value problem, since the exact boundary is not known. We are interested in the relationship between the average scattered field or radar cross section and the average surface properties. The radar cross section normalized by the area A, defined as,

$$\gamma(\theta_i) = \gamma(\theta, \theta, \phi) \tag{1}$$

θ_i, θ_s, and ϕ_s are shown for the bistatic case in figure 1 below. The rough surface is assumed to consist of height variations about a mean plane, which is taken as the xy-plane. The incident wave line in the sz-plane at a polar angle.
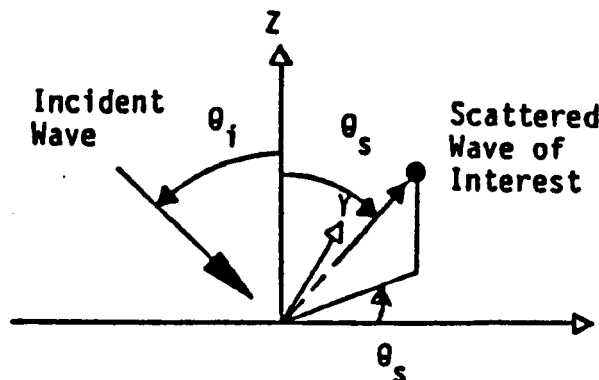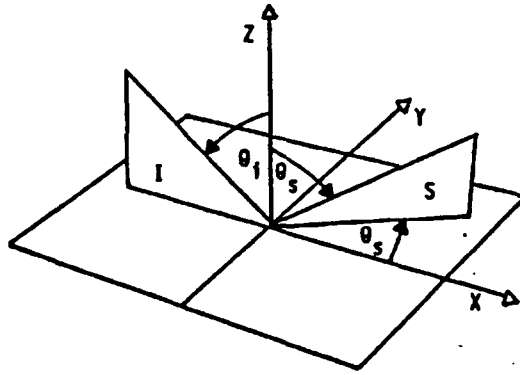

Figure 1 - THE "BISTATIC SCATTERING GEOMETRY"

Figure 2- THE "SCATTERING GEOMETRY"

Figure 1 shows the bistatic scattering for a planar rough surface whose mean height coincides with the xy-plane and figure 2 shows the scattering geometry. I is the plane of incidence and S is the scattering plane.

The symbol $<\quad>$ refers to an average and the normalized average backscattering cross section are defined as

$$\gamma(\theta_i) = \gamma(\theta, \theta, \phi) \tag{2}$$

Usually one has a specific purpose or reason for investigating scattering from a rough surface, these goals may be divided into three categories: (1) The problem of direct scattering; one wishes to know the average properties of the scattered signal or cross section when the surface properties of rough surface are known, and the scattering information is expressed in terms of the surface properties. (2) The problem of inverse scattering; one wishes to obtain statistical information about the rough surfaces from a knowledge of the average properties of the scattered field. This problem is more difficult in that there appear to be many classes of rough surfaces producing the same average scattering cross section as a function of the bistatic scattering angle and wavelength.

The radar cross section (scattered field) can be expressed as 4π times the power delivered per unit solid angle in the direction of the receiver divided by the power per unit area incident at the target. The factor 4π enters from the definition of a solid angle. Assuming for the moment that the propagation path between the target and the receiving system is lossless, then this power ratio may be expressed as

$$4\pi r^2 \frac{(E^s \cdot E^{s*})}{(E^i \cdot E^{i*})} = 4\pi r^2 \frac{(H^s \cdot H^{s*})}{(H^i \cdot H^{i*})}. \tag{3}$$

Where E^S and H^S are the scattered electric and magnetic fields, respectively and E^i and H^i are the incident fields. The scattered field is defined to be the difference between the total field (with the target present) and the incident field. This is summarized as

$$E^S = E^T - E^i. \tag{4}$$

Finding the cross section (σ) now becomes a problem in electromagnetic field theory. In order that the cross section be independent of r, it is desirable to let r in Eq. 3 become arbitrarily large (scalar definition, not a tensor function).

$$\sigma = 4\pi \lim_{r \to \infty} r^2 \frac{(E^s \cdot E^{s*})}{(E^i E^{i*})} = 4\pi \lim_{r \to \infty} r^2 \frac{(H^s \cdot H^{s*})}{(H^i \cdot H^{i*})}. \tag{5}$$

To compute the scattered field, one sometimes computes the current induced on the target and then treats the target current distribution in terms of an equivalent aperture distribution. Antennae are often compared to an isotropic antenna, that is, anantenna which radiates uniformly in all directions.

In the problem of clutter, returns from terrain are not wanted. However,in many cases the presence of clutter is unavoidable along with the desired signal. One can detect and analyze the desired signal significantly better if more is known about the properties of the clutter or noise produced by terrain scattering. It is noted that surface information is generally known, and the properties of the scattered signal are related to Eq. 3.

The Fourier transform was used in this analysis and it was shown that space-time equations can be readily derived from their space-frequency counterparts. The equations derived will be applicable to field of arbitrary time variation.

The source signal is related to the fourier transform, F(w), by:

$$(a) \quad F(w) = \int_{-\infty}^{\infty} f(t)e^{jwt}dt \qquad (b) \quad f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(w)e^{-jwt}dw. \tag{6}$$

The upper case lettering are used for the transformed functions unless otherwise noted. The wave number k = w/c, where c is the speed of light in the medium. α and β are the direction cosines of the vector r projected onto the xy-plane with $\alpha = \sin\theta\cos\theta$, $\beta = \sin\theta\sin\theta$.

It can be assumed that the field distribution over the antenna apertures and object plane is space-time separable; the field strength of each radiating element observed at some point r is e(r,t) = e(r) f(t - r/c), the field strength, and f(t - r/c) is the sine wave traveling at the speed of light.

Assume the scatter in figure 2 to be placed in a far field, such that the assumption of uniform impinging planewave over the object can be obtained by the application of the far field Kirchhoff-Huygens principle.

The basic principle of time-domain imaging over an object plane has been formulated and discussed by many researchers. It was not until 1983 when Wolfgang-M Boerner developed a finite difference-time domain method (FD-TD) with a near field to far field transformation using field equivalence, that a method to treat realistic scattering problems effectively was developed. In this method, the scattering problem is analyzed in two steps by treating the relatively complex near field region and the relatively simple far field region separately. The method involves first the determination of equivalence electric and/or magnetic current tangential to a virtual surface surrounding the scatter of interest by using the FD-TD method for a given external illumination. The computed near-field equivalent currents are then transformed to derive the far-field scattering problem and the radar cross section. Since the FD-TD method can deal with dielectric, permeable, and in homogeneous

materials in a natural manner, it is possible to incorporate most of the physics of wave interaction with any complex scatter of interest

The Kirchhoff-Huygens principle deals with scattering, but the process is not as sensitive in analyzing scattering as the method put forth by Wolfgang-M Boerner, et al. TheKirchhoff- Huygens method deals with far field scattering, whereas Wolfgang- M Boerner treats both far field and near field scattering.

The theory involving finite difference-time domain (FD-TD) is not without limitation. This method is based upon the physical optics approximation which is valid for scatter whose dimension is significantly larger than the incident wave length and is independent of polarization.
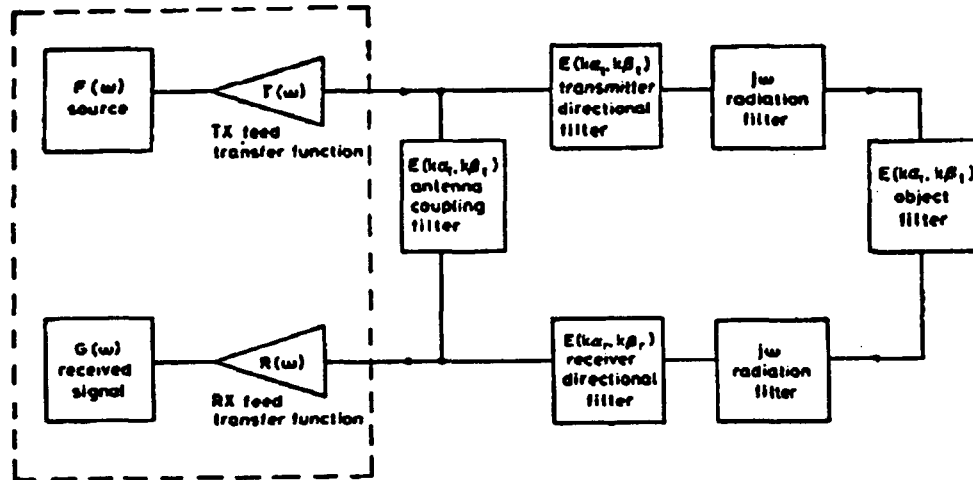

Figure 3 - SCHEMATIC DIAGRAM OF THE TARGET IMAGING SYSTEM

$$I(w) = \frac{\cos\theta}{4\pi c r_o} jw F(w) T(w) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e(x,y) exp[-jk(\gamma_o - \alpha_t x - \beta_t y)] dx dy. \tag{7}$$

$$I(w) = \frac{\cos\theta}{4\pi c r_o} jw F(w) T(w) E_t(k\alpha_t, k\beta_t) exp[-jkr_o]. \tag{8}$$

Where $E_t(k\alpha_t, k\beta_t)$ is the transmitting-antenna anisotropic directional filter and r is the distance between the transmitter aperture and an object plane taken as reference.

The second application of the Kirchhoff-Huygens principle and fixed phase evaluation can be shown with little calculus that the received waveform, g(t), is

$$g(t) = const \int_{-\infty}^{\infty} F(w) T(w) R(w) [E_a(k\alpha_o, k\beta_o) + (jw)^2 E_t(k\alpha_t, k\beta_t) E_r(k\alpha_r, k\beta_r) E_o(k\alpha_o, k\beta_o) e^{-jkr}] e^{-jwt} dw. \tag{9}$$

where the direction cosine, the constant and the distance are dependent on the imaging geometry.

For the monostatic case commonly encountered (see figure 4 below), the transmitter and receiver coincide. In particular, if the phase center of the object is on boresight which suggests that the target is closely tracked, Eqn. 9 can be considerably simplified to yield:

$$g(t) = const \int_{-\infty}^{\infty} F(w) T(w) R(w) [E_a(k\alpha_o, k\beta_o) + (jw)^2 E_t(0,0) E_r(0,0) E_o(k\alpha_o, k\beta_o) e^{-2jkr_o}] e^{jwt} dw. \tag{10}$$

The delineation of the objective response is exercised as follows: First, the time response in the absence of any scattering objects is recorded.
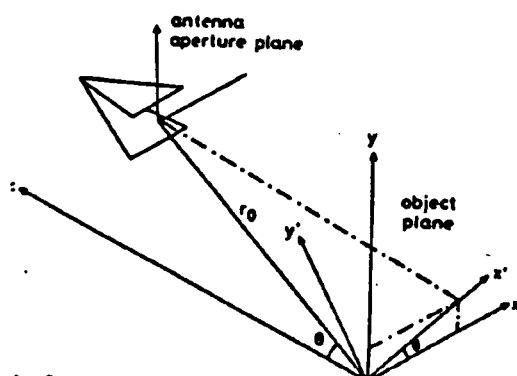

Figure 4 - IMAGING GEOMETRY

$(a) \quad z = \dfrac{\alpha z' + \beta y'}{\sqrt{\alpha^2 + \beta^2}}$ $(b) \quad y = \dfrac{\beta z' - \alpha y'}{\sqrt{\alpha^2 + \beta^2}}$

$(c) \quad \cos\theta = \dfrac{\alpha}{\sqrt{\alpha^2 + \beta^2}}$ $(d) \quad \sin\theta = \dfrac{\beta}{\sqrt{\beta^2 + \alpha^2}}$
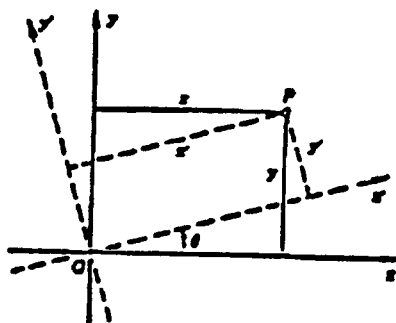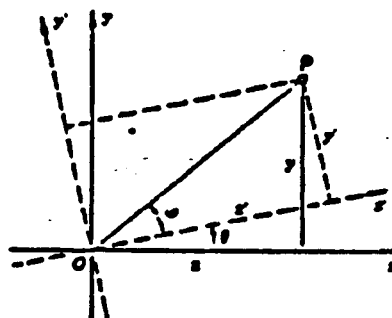
IMAGING GEOMETRY EQUATIONS

Figure 5a

Figure 5b

$(f) \quad z = z'\cos\theta - y'\sin\theta,$ $(g) \quad z' = z\cos\theta \div y\sin\theta,$

$\qquad y = z'\sin\theta \div y'\cos\theta,$ $\qquad y' = -z\sin\theta \div y\cos\theta.$

$(h) \quad \dfrac{z}{|OP|} = \cos(\phi \div \theta), \qquad z = |OP|\cos(\phi \div \theta) = |OP|(\cos\phi\cos\phi - \sin\theta\sin\theta).$

$(i) \quad \sin\phi = \dfrac{y'}{|OP|} \quad and \quad \cos\phi = \dfrac{z'}{|OP|}$

THE ROTATION TRANSFORMATION

The antenna's coupling filter is removed when the signal is subtracted from other records with the scatter present. Multiple path scattering between the object and surrounding obstacles can be gaited out in time when sufficiently large space is given. By definition,

$$E_o(k\alpha_o, k\beta_o) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e_o(x,y)exp[2jk(\alpha_o x + \beta_o y)]dxdy. \tag{11}$$

The system kernel, K(w), can be obtained once the coupling term has been eliminated.

$$K(w) = cost(jw)^2 F(w)T(w)R(w)E_i(0,0)E_r(0,0)E_o(0,0)e^{-2jkr_o}. \tag{12}$$

The use of mathematical techniques, provides an opportunity for the backscattering waveform to be simplified. To give the system kernel, K(w), and to obtain the electrical field strength in frequency domain and by the projection theorem, the response obtained is equivalent to a slice in the two dimension fourier plane (figure 4), which is a representation
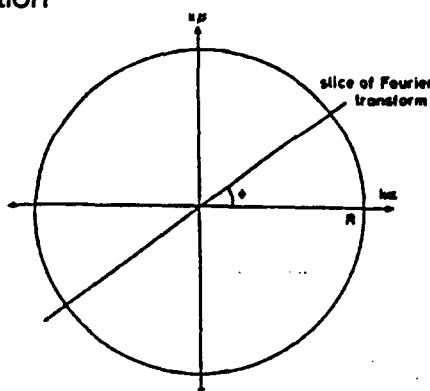

Figure 4 - ONE-DIMENSIONAL TARGET FREQUENCY RESPONSE

of the one-dimensional target frequency response in the two- dimensional fourier plane at a particular viewing direction. Equation 9 becomes

$$E(k\alpha, k\beta) = \frac{G(w)e^{2jkr_o}}{k(w)} \tag{13}$$

The application of the inverse fourier transform of equation 13 yields the object impulse response

$$\int_{-\infty}^{\infty} E(k,o)e^{-jw(t-2x'\sin\frac{\theta}{c})}dw = e(t - 2x'\sin\frac{\theta}{c}) \tag{14}$$

which can be considered as a series of scans across the object surface where each is an integral of the object field which is orthogonal to the viewing direction. The spatial resolution is thus achieved by the utilization of the time channel and the minimal resolvable dimension Δx is related to the sampling integral T, namely

$$\Delta x = \frac{cT}{2sin(\theta)} \tag{15}$$

a different impulse response is obtained by the rotation of the object plane about the z-axis, which corresponds to another slice in the Fourier plane. Further rotation over a 180 degrees span will fill up the entire Fourier plane.

The two-dimension object field distribution is recovered, namely the recnstruction strength, e(x, y)

$$e(x,y) = \frac{1}{4\pi^2}\int_0^\pi d\phi \int_{-R}^R \frac{R}{\pi}E(k\alpha,k\beta)\sin(\frac{k}{R}\pi)e^{-2jk(\alpha x+\beta y)}dk \tag{16}$$

a smoothing filter A($k\alpha,k\beta$) is introduced to compensate for the amplification of high frequency component due to the factor k in field strength equation [Eq. 16]. It is not necessary to have A, for noise free data. For practical numerical evaluation, the finite sampling rate will impose an upper limit, R, to the wave number k. To eliminate abrupt cutoff, which would cause ringing in the resultant image, it is necessary to introduce equation 16, to yield a smoothed image, equation 17

$$A(k\alpha,k\beta) = \sin\frac{(k\pi R)}{(\frac{k\pi}{R})} \tag{17}$$

$$e(x,y) = \frac{1}{4\pi^2}\int_0^\pi d\phi \int_{-R}^R \frac{R}{\pi}E(k\alpha,k\beta)\sin(\frac{k}{R}\pi)e^{2jk(\alpha x+\beta y)}dk \tag{18}$$

and is the basis of our reconstruction field strength.

III. Experimental

The government, private, and public sector have always had a need for time-domain imaging. This area of research did not surface until early 1962. The new and old researchers in electromagnetic field theory and other scientist provided many theoretical ideals on direct and indirect scattering. It was observed that these researchers did not have an insight into the circuit design, arrangement of apparatus and development of the necessary experiments to verify and determine the applications of the existing theory.

Back scattering from metallic objects was experimentally observed and the signal processing technique for extracting the object response was performed in 1985 by Yeung and Evans. It was not until this time when Yeung and Evans put the existing theory coupled with experimental verification together and made a projection of the long over due applications. After carefully reviewing the literature and the experimental verification provided by these two researchers, it was observed that applications of time-domain imaging is applicable to many aspects of the private and public sector, but the immediate and long range beneficiaries are NASA and the military.

Recently, Dr. Kumar Krishen of NASA/JSC has proposed the use of time-domain imaging for space robotic vision application. It has been shown that space objects are covered with heat shields, (dielectric materials), and microwave penetrate through these shields. A multi-sensor approach to this robotic vision application has been shown to have several advantages over a video approach.

The authors of this paper will use the necessary apparatus and experimental arrangement of Yeung and Evans to demonstrate the viability of impulse imaging. Verifying the experimental results should give us confidence in the validity of the theoretical method over a broad range of angles and the enormous experimental

potential of time-domain techniques in radar- scattering measurements. After scrutinizing the theory and experimental results of Yeung and Evans, the authors of this paper will look into time-domain robotic vision application and the effects of pulse width, polarization, look angles, and phase images of robotic vision coupled with, conducting experiments to verify the application of the existing theory and experiment.

The target size employed in these experiments are comparable to the source pulse width, and it is well known that electrical shape of objects at low frequencies could be considerably different from the physical shape.

Finite source rise-time (hence limited bandwidth) and insufficient source power are the major difficulties for achieving the highest resolution and sharp images. The system kernel involves the differentiation of the source waveform, and this adds uncertainties in the determination of low-frequency components which is usually manifested as a slowly varying baseline in the deconvolved result.

IV. Applications

In 1986, a summary of goals and objectives were recommended by the NASA's Advanced Technology Advisory Committee (ATAC) [1]. The robotic applications set up by the Committee are: (a) Teleoperation of mobile remote manipulator with collision avoidance. (b) Mobile multiple-arm robot with dexterous manipulators to inspect and exchange orbital replaceable units. (c) Systems designed to be serviced, maintained, and repaired by robots.

The current remote manipulator system (RMS) will be used to manipulate initial station assembly in the Space Shuttle. Further versions of the RMS will be employed to assemble Space Station and satellite.

NASA/JSC is currently developing an early robot, the extra- vehicular astronaut (EVA) retriever, to perform as an aid to an EVA/man maneuverable unit (MMU) astronaut and retrieve the astronaut if circumstances dictate such an operation.

In the development of the space vision systems cost effectiveness, speed, small size, lightweight, high reliability and flexibility, and ease of operation must be considered. Microwave techniques of time-domain imaging is one of the tools to achieve the above purposes.

Robotic technology has been widely utilized by NASA to explore space; however, one of the most challenging problems to deal with is robotic vision. Vision is needed in fixed locations to keep track of location of objects, to pick and place machines/systems, and to monitor performance of non-automatic systems. Indeed, vision provides intelligence and flexibility in locating automated and non-automated machines.

As the robotics era dawns in space, vision will provide the key sensory data needed for multi-faceted intelligent operations. In general, the 3D scene/object description along with location, orientation, and motion parameters will be needed. Sensor complements may include both active and passive microwave and optical types with multi-function capability. The fusion of the information from these sensors to provide accurate parameters for robots provides by far the greatest challenge in vision. Furthermore, the compression, storage, and transmission of the

information associated with multi-sensor capability require novel algorithms and hardware for efficient operation.

The vision requirements for space robotics are characterized by environmental factors and tasks the robot has to perform. The natural space environment consists of intense light and dark periods. At a normal Space Station altitude, the sunlight intensity will fluctuate above 60 minutes of extreme brightness and 30 minutes of darkness. Furthermore, due to the absence of atmosphere, light is not diffused/scattered. The ubiquity of white surfaces intensifies the problem of relying on photometric data for object identification/discrimination. A secondary source of concern effecting vision is the absence of gravity. For free flying and tethered objects may be found due to the lack of disturbance caused by aerodynamic and gravitational forces.

The unpredictable nature of maintenance and repair tasks creates a problem in the development of the capability/design of space robots. The vision capabilities must be adaptive/versatile to accommodate these uncertainties

The quest for the highest resolution microwave imaging and the principle of time-domain imaging has been the primary motivation for recent developments in time-domain techniques. In the last decade, modern techniques in sampling devices and advent fast pulse generators has brought new technology to the practicing radar engineers in the measurements of picoseconds at a greatly reduced cost. With the present technology, fast time varying signals can now be measured and recorded both in magnitude and in-phase.

Conventional radar ranging has been enhanced by modern technology's ability to extract details about the scattering object.

V. Conclusions

In this paper various research approaches to time-domain imaging have been reviewed with particular emphasis on the mathematical and physical approximations made in them.

Yeung and Evan dealt with the Kirchhoff-Huygens Principles, which deal with far field scattering whereas Boerner dealt with both far field and near field scattering. We feel that the far and near field scattering technique is a better approximation, particularly when applied to a scattering object that is not a perfect conductor.

Based on the theoretical and experimental verification provided by these authors it is necessary that future research be done so as to go beyond the existing approximations of Boerner, et al. For instance, a non-trivial generalization, should apply to conductivity and non-conducting targets.

Acknowledgement

Bibliography

Advancing Automation and Robotics Technology for the Space Station and for the U.S. Economy, Progress Report 3. NASA Advanced Technology Advisory Committee, Washington, D.C., September, 1986.

Space Station Automation Study, Automation Requirements Derived from Space Manufacturing Concepts, Volume I and II, General Electric, November 27, 1984

Automation Study for Space Station Subsystems and Mission Ground Support-Final Report, Hughes Aircraft Company, November 1984.

Space Station Automation Study, Final Report, Volume I and Volume II, Martin Marietta, November 1984.

Space Station Automation and Robotics Study, Final Report, Boeing Aerospace Company, November 1984.

Satellite Services System Analysis Study, Executive Summary, Final Briefing, Grumman Aerospace Corporation, July 22, 1981.

NASA Space Station Automation: AI Based Technology Review, Executive Summary, SRI International, March 1985.

Bachman, C.G., *Radar Targets*(Gover Pub. Co. Ltd., 1982), pp. 110-113.

Beckman, P. and Spizzichino, A., *The Scattering of Electromagnetic Waves from Rough Surfaces*.(The Macmillan Co., 1963), pp. 10-28, 178-181.

Bennett, C.L., and Ross, G.F., *Time domain electromagnetics and its applications*, ibid., 1978, 66, pp. 299-318.

Boerner, W.M., Ho, C.M., and Foo, B.Y., *Use of Radon's projection theory in electromagnetic inverse scattering*, IEEE Trans., 1981, AP-29, pp. 336-341.

Chan, C.K., and Farhat, N.H., *Frequency swept tomographic imaging of three dimensional perfectly conductiong objects*, ibid, 1981, AP-29, pp. 312-319.

Censor, Y., *Finite series-expansion reconstruction methods*, ibid, 1983, 71, pp. 409-419.

Evans, S., and Kong, F.N., *Gain and effective area of impulse antenna, Third International Conference on Antenna and Propagation*, ICAP 83, Norwich, England, april 1983, pp. 421-424.

Fialkovskiy, A.T., *Diffraction of planar electromagnetic waves by a slot and a strip*, Radio Eng. Electron., 1966, 11, pp. 150-157.

Kennaugh, E.M. and Moffatt, D.C., *Transient and impulse approzimations*, Proc. IEEE, 1965, 53, pp. 893-901.

Kong, F.N., Ph.D. Dissertation, Cambridge University, August 1983, Chap. 8.

Lewitt, R.M., *Reconstruction algorithms: transform methods*, Proc. IEEE, 1983, 71, pp. 390-408.

Reader, H.C., Evans, S., and `Yeung, W.K., *Illumination of a rectangular slot radiator over a 3 Octave bandwidth*, Fourth International Conference on Antennas and Propagat., ICAP 85, 1985, pp. 223-226.

Ruck, G.T., Barrick, D.E., Stuart, W.D., and Krichbaum, C.K., *Radar Cross Section Handbook*(Plenum Press, 1970), pp. 671-689.Schubert, K. A., Young, J.D., and Moffatt, D.L., *Synthetic Radar Imaging*, ibid, 1977, AP-25, pp. 477-483.

Tiknonov, A.N. and Arsenic, V.Y., *Solutions of ill-posed problem*(Winton-Wiley, New York, 1977)

Yeung, W.K. and Evans, S., *Time-domain microwave target imaging*, Proc. IEEE, 1985, 132, pp. 345-350.

ARC JET DIAGNOSTICS TESTS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1988

Johnson Space Center

Prepared by: Ronald J. Willey, Ph.D.

Academic Rank: Assistant Professor

University & Department: Northeastern University
Department of Chemical Engineering
Boston, Massachusetts 02115

NASA/JSC

Directorate: Engineering

Division: Structures and Mechanics

Branch: Thermal

JSC Colleague: John E. Grimaud

Date Submitted: August 19, 1988

Contract Number: NGT 44-005-803

ABSTRACT[1]

Two objectives were addressed during a 10 week 1988 NASA/ASEE summer faculty fellowship at the Johnson Space Center Atmospheric Reentry Materials Structures Evaluation Facility (ARMSEF). These objectives were the evaluation of mass spectrometry for the measurement of atomic and molecular species in an arc jet environment, and the determination of atomic recombination coefficients for RCG (reaction cured glass) coated HRSI (high temperature surface insulation) materials subjected to simulated reentry conditions.

Evaluation of mass spectrometry for the measurement of atomic and molecular species provided some of the first measurements of point compositions in an arc jet tunnel environments. A major objective of this project centered around the sampling residence time. A three staged vacuum sampling system pulled the molecules and atoms from the arc jet to a quadrupole ionization mass spectrometer in 400 milliseconds. Conditions investigated included a composition survey across the nozzle exit at 3 cm z-distance from the nozzle exit for 3 different currents. Also, a point composition survey was taken around a shock created by the the presence of a blunt body.

Results for the composition survey across the nozzle exit showed concentration gradients for oxygen and its related secondary species: oxygen atoms, nitric oxide, and carbon dioxide. This gradient was more pronounced at higher currents and locations closer to the nozzle exit. The gradients were skewed from the centerline with maximum concentration occurring at a theta of 1.0 degree. A composition survey across the shock showed nitric oxide concentrations dropped while water concentrations increased about 6 fold. Other species also showed small changes. By using the existing copper pitot tube, little disturbance occurred to the arc jet operations.

Recombination measurements on RCG coated HRSI materials showed low catalytic recombination coefficients for both nitrogen and oxygen atoms. The experiments were performed on a 5.08 cm radius HRSI hemispherical model (LI-900) coated with a reaction cured glass coating. Gas mixtures investigated included pure nitrogen, 11%, 14% and 23% oxygen in nitrogen. Flowrates and currents ranged from 0.036 to 0.056 kg/sec, and 300 to 550 amps respectively. The results showed that the nitrogen recombination coefficient is relatively constant at 0.014 over a temperature range of 1350 to 1550 K. These results compare favorably with those reported by Scott (1981) and Kolodziej and Stewart (1987). The analyses of air runs were not as conclusive. Experimental repeats within a run were good, however, repeats outside of a run showed a variation in surface temperatures of 60 K. Oxygen concentration appears to have an influence, however, more measurements are recommended.

---

[1]This report is dedicated to Dr. Frederic A. Wierum - a pioneer in Arc Jet Diagnostics

INTRODUCTION

This report is divided into two major sections. The first section will discuss mass spectrometry of atomic and molecular species in an arc jet. The second section will discuss atomic recombination measurements on RCG coated HRSI materials.

SECTION 1

EVALUATION OF MASS SPECTROMETRY FOR ARC JET SPECIES IDENTIFICATION

BACKGROUND

Quadrupole ionization mass spectrometry is a well developed analytical technique used for the determination of gas compositions via mass separation. The principle centers around an ionization filament which emits electrons. These free electrons strike a fraction of molecules and atoms as they pass through the ionization cage. These molecules and atoms lose an electron as a result of this collision and a charged positive ion results. Under the presence of a rf voltage superimposed on a dc voltage established by a two sets of oppositely charged rods (quadrupoles), certain charged ions can pass through the quadrupoles and strike a detector (a faraday cup in this instance). By ramping a potential across the quadrupoles a "scan" can be performed which will measure the m/e (mass to charge ratio) present in the mass spectrometer at any given time. Ions created in the process can also fragment and thus a parent molecule like a nitrogen will have a signal at 14 and 28 plus a small amount at 29 for the NN^{15} isotope. Further details about mass spectrometry may be found in a book by McLafferty.

Mass spectrometry has been attempted in arc jet environments before with limited success. A major limitation was residence time. Real time responses were on the order of 30 minutes – a response time much too long. A mass spectrometer sampling system developed by R.J.Willey at Northeastern University has enable rapid response of an order of 1 second for gas samples from an atmospheric environment. This rapid response was achieved by a staged vacuum system. Therefore an earlier drawback could be eliminated if the sampling system could be adapted to an arc jet environment.

Another concern in mass spectrometry as applied to arc jet environments is the insurance that what is measured is what appears in the stream. The presence of high atom concentrations in the stream present a possible sampling problem because of atomic recombination in the probe and sample line. Recombination reactions can occur in both the gas phase and along the probe walls as the gas sample travels from the source to the mass spectrometer. A preliminary study completed at Northeastern University showed that homogeneous gas phase recombination reactions could be virtually eliminated by the running at pressures below 20 microns at a residence time of 1 second. Wall reactions could be eliminated by using a specially designed probe constructed an inert wall material (quartz).

EXPERIMENTAL

A block diagram of the mass spectrometer system is shown in Figure 1. The initial study used the existing pitot tube (constructed of copper) as the sample probe. The pitot tube was tied to the mass spectrometer by a 2 m 0.188 I.D. Resistoflex tubing (Teflon^{TM} lined) and 1.5 m 1/4" O.D. Eastman Nylon^{TM} tubing. Four valves separated 3 vacuum stages and the mass spectrometer chamber. These valves were a solenoid valve which allowed remote sampling capability, 2 Nupro SS-4H bellow vacuum valves and one Nupro SS-4BG-TW bellow vacuum valve. The vacuum pumps were standard mechanical vacuum pumps manufactured by Sargent Welch, Duo Seal and Cenco. The mass spectrometer chamber was evacuated by a Balzer's turbo molecular pump Model # TPH050 and this pump was backed by an Edward's Model E2M-1 high vacuum pump.

The mass spectrometer used was an UTI AMX-100. The mass spectrometer has a m/e range of 1 to 100. Thus all light gases below a molecular weight of 100 can be observed. Signals from the mass spectrometer were interfaced to an IBM-XT computer through a Cybord A/D and D/A converter (Isaac Model 91-I). Programming written at Northeastern University allowed for the collection of mass scans or selected peak scanning as a function of time.

Experimental conditions investigated were:

Currents: 400, 800, 1100 amps

Pitot z-distances from the nozzle: 3 cm, 23 cm, 43 cm

Radial Positions from the centerline: +/- 7.9, 6.32, 4.74, 3.16, 1.58, 0 degrees

Gases: Air, Nitrogen, and Argon

RESULTS AND DISCUSSION

Shown in Figure 2 is the mass spectrometer response at m/e= 30 (Nitric Oxide) for a composition survey at 3 current levels across an arc normal to the nozzle exit plane and located 3 cm from the nozzle exit. Several phenomena are evident. As the current is increased, nitric oxide concentrations increase, in this case from 2% to 11% at the centerline. Further, the results showed more nitric oxide at the centerline then at the edges of the nozzle thus indicating a concentration gradient across the nozzle. Gradients were also observed for oxygen molecules, oxygen atoms, and carbon dioxide.

Also shown in Figure 2 is a pitot pressure survey. Note that it does not show as high as % maxima. Thus the peaks seen at 800 and 1100 amps for nitric oxide suggest that centerline concentrations for this species are higher. Another source for nitric oxide may be the reaction of nitrogen atoms with oxygen atoms in the probe. However as is discussed below, nitrogen atoms are reaching the mass spectrometer and it is currently suspected that the nitric oxide measured is coming from the arc tunnel. The skewness shown in the figure may represent an arc column slightly out of line.
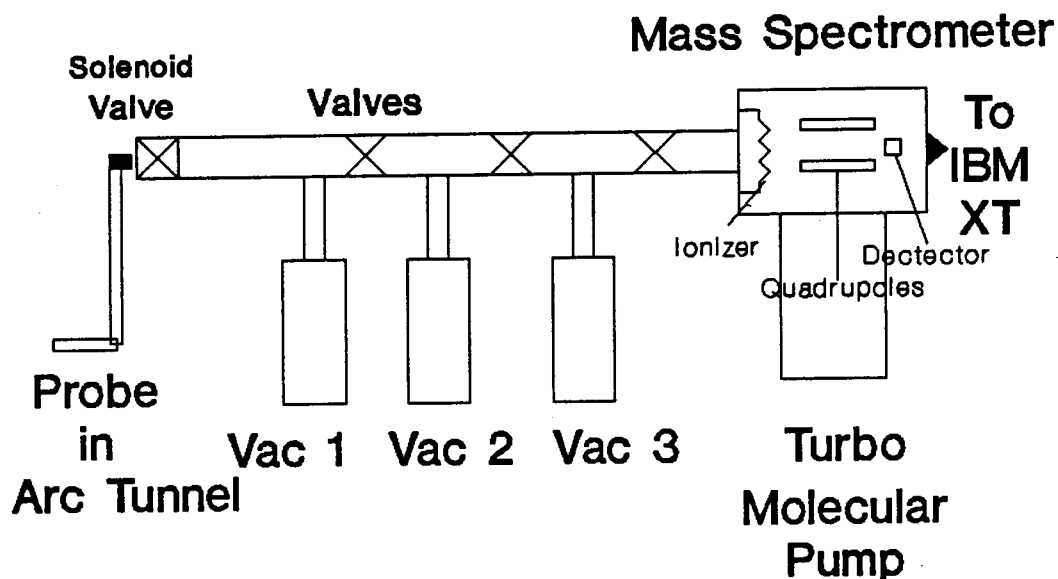
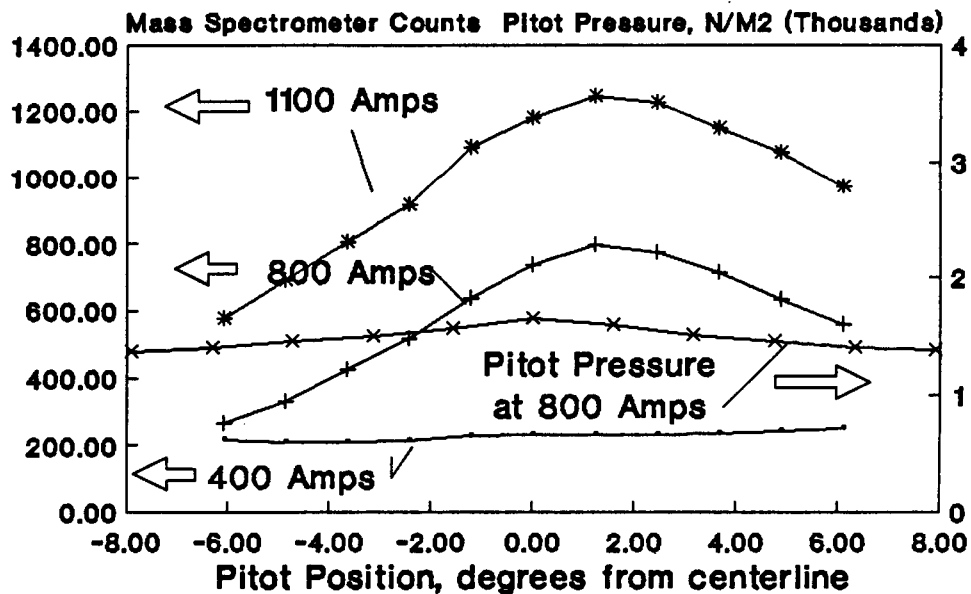Figure 1. Block Diagram for the Mass Spectrometer System.

Figure 2. Example of Nitric Oxide Concentration for Composition Survey across the Nozzle Exit Conducted at Three Currents.

Figure 3 shows the corrected mass spectral response at m/e = 14 (N), 16 (O), 28 (N_2), 30 (NO), and 32 (O_2). Corrected response means that the signals for each of the m/e ratios have had parent molecule fragments subtracted. For example, the m/e signal at 14 will have a contributions from N_2 and NO. Thus the 14 signal was corrected by subtracting the contribution due to N_2 and NO. Note in this figure how O_2 trends drop slightly as current is increased, however, it is a significant response. The source of the O_2 is suspected to be from wall recombination of oxygen atoms. Using a wall recombination coefficient of oxygen of 0.06 for copper lined sections of the probe system and 0.0005 for the inert wall probe areas results in an approximation that 70% of the oxygen atoms may be lost to wall recombination. It is estimated that this figure can be lowered to 30% by simply using the quartz sample probe.

High nitrogen atom concentrations are somewhat puzzling. The homogeneous reaction models predicted that in the presence of O atoms, N atoms would convert to nitric oxide in the probe. However, the m/e signal at 14 is significant at about 14% of the total mass spectral response even after the signal is adjusted for fragments from N_2. Thus, nitrogen atoms do reach the mass spectrometer.

An indication of nitric oxide concentrations around a shock is shown in Figure 4. Figure 4 shows the nitric oxide concentration decreasing dramatically inside a shock. An explanation for this observation is that nitric oxide dissociates across the shock to nitrogen and oxygen atoms. If the source of nitric oxide was probe reactions, one would expect that nitric oxide concentration would increase across a shock. Spectrographic work by Willey (1987) showed the primary emitter in the free stream to be nitric oxide. Thus conditions may exist where nitric oxide concentrations are above 1% in the arc jet. Mappings for other species showed the definite shock boundary, however, their concentrations were relatively constant on both sides of the shock with the exception of water. Water increased about 4 to 6 fold on the backside of the shock. This result suggests that some moisture is being pulled in from the ejector system. Further information about this project will be available in a separate report.

CONCLUSIONS

1. Sampling times of less than a second can be achieved.

2. Nitric oxide is a good probe molecule.

3. Concentration gradients occur across the nozzle and are strongest for nitric oxide.

RECOMMENDATIONS

1. Install the quartz sample probe and 1/2" teflon or nylon lines throughout the sampling system so a non-catalytic sampling system can be compared to the existing pitot tube sampling system.

2. Verify the presence and percent concentrations of nitric oxide observed at higher currents spectrographically.

3. Install pyrex orifices in place of valves and use inert valve bodies for locations where valves are necessary (solenoid valve and mass spect chamber valve).
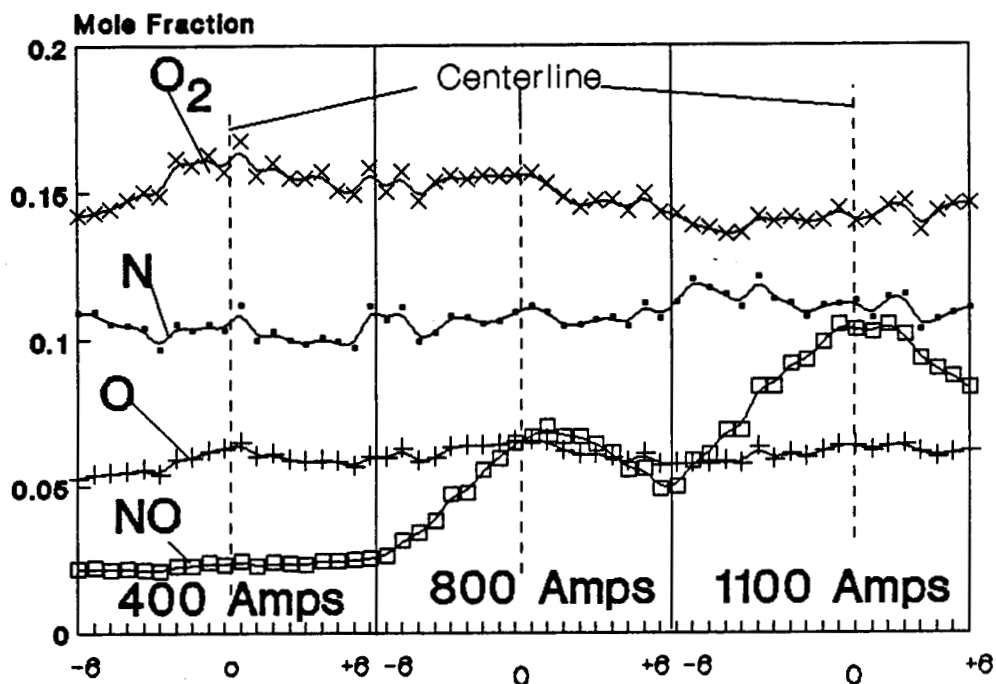
Figure 3. Mole Fractions of Oxygen, Oxygen Atoms, Nitrogen Atoms, and Nitric Oxide as a Function of Pitot Position and Current.
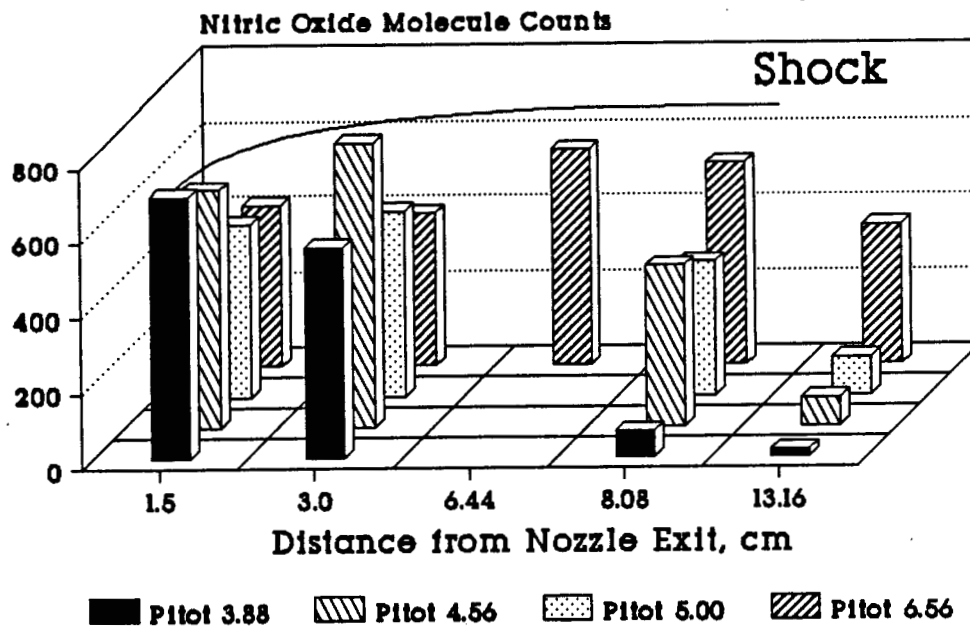

Figure 4. Mass Spectrometer Signal Counts at m/e=30 (NO) for a Shock Survey around a Blunt Body.

SECTION 2

CATALYTIC MEASUREMENTS ON HRSI AND RCC MATERIALS

BACKGROUND

This objective involved the determination of atomic recombination rates on HRSI (high temperature reuseable surface insulation) and RCC (reinforced carbon carbon) materials at various gas compositions, flow rates, and enthalpies.

Knowledge of catalytic responses on thermal protection systems are of tremendous importance in heat shield design. The difference between a non catalytic surface and fully catalytic surface influences the rate of heat transfer to the surface during a reentry manuever, and therefore, the difference in temperature measured at the surface. The ideal surface would be non catalytic i.e. atoms would strike the surface and reflect off. On the other hand, on a fully catalytic surface all atoms which strike the surface recombine to the ground state, and the energy released in this process would be transferred to the surface. Thus the surface can heat up tremendously. Demonstration of the catalytic effect on surfaces was reported for STS-2 which had 2 catalytic points on its thermal protection system (Stewart, 1981) and subsequent analysis predicted surface temperature increases of 270 K at these catalytic points (Scott, 1982).

Data analysis followed an approached developed by Scott (1981,83). The analysis begins by using simple flow conservation equations for an inviscid stream tube and frozen flow relations across the shock wave (Scott, 1983). Shock properties are determined using the measured quantities of enthalpy, stagnation pressure, and mass flowrate along with the calculated effective area taking into account the boundary-layer effect. The effective area was determined using the NATA code (Bade and Yos, 1976). Then the Fay and Riddell (1958) relation as modified by Goulard (1958) and further refined by Scott (1981) was used to generated curves for the ratio of heating rate to fully catalytic heating rate versus recombination coefficient. Since recombination can occur for both nitrogen and oxygen atoms the approach requires separating the influence of nitrogen before analyzing results for air. This is done by running tests in pure nitrogen conditions and obtaining the nitrogen recombination coefficient in terms an Arrhenius expression which would serve as an input into the air analysis. Further details about the method can be found in Scott (1981, 83).

EXPERIMENTAL

Hemispherical test articles made out of HRSI (LI-900) RCG (reaction coated glass) were place in the 5-MW arc jet chamber at JSC. The arc column consisted of 5 heater packs attached to a 2.25 inch throat which was connected to a 15° solid angle nozzle with an exit diameter of 15 inches. Further details about the facility made be found in Rochelle et al. (1983). The test articles were placed in the right arm while a comparable metallic hemispherical model instrumented with slug calorimeters was placed in the left. The slug calorimeter served as a reference heating rate measurement. The heating rate was determined by measuring the rise in

temperature divided by the change in time during the period the calorimeter was inserted into the stream. The constant for the calorimeter was 0.930 BTU/ft^2/$^{\circ}$F The recombination coefficients for the calorimeter were assumed constant at 0.6 for nitrogen and 0.15 for oxygen.

The procedure was:

1. Arc on.

2. Run for 2 to 5 minute period to allow gas flows and arc conditions to stabilize.

3. Swing in the left arm (calorimeter) into the stream for a period of 1 to 2 seconds.

4. Swing in the right arm (the model) into the stream wait until the stagnation point thermocouple reads constant, take a test point.

5. Swing in the pitot tube into the stream, take a test point to obtain the stagnation pressure.

6. Change to the next current setting.

7. Repeat steps 3 to 6 until full ranges have been covered.

The range of conditions studied were:

Nitrogen	0.038 kg/sec (0.08 lb/sec)	at 400 to 550 amps
11% (mass) O_2 in N_2	0.038 kg/sec (0.08 lb/sec)	at 300 to 400 amps
14% O_2 in N_2	0.056 kg/sec (0.14 lb/sec)	at 300 to 340 amps
23% O_2 in N_2 (Air)	0.038 kg/sec (0.08 lb/sec)	at 350 to 450 amps

The sequence of testing was:

7/12/88   Air - 12 points followed by Nitrogen - 1 point

7/14/88   Nitrogen - 7 points followed by 14% O_2 - 3 points

7/15/88   Air - 3 points followed by 11% O_2 -6 points followed by N_2 - 3 points

The experimental procedure included several repeat points in order to check repeatability of the results.

RESULTS AND DISCUSSION

Table 1 presents conditions tested and resultant heating rates and temperatures measured. Figure 5 shows that good repeatability could be achieved within a run, however, between runs temperatures for air conditions varied by about 60 K. Another observation made is that in the presence of 11% oxygen, the model temperatures were hotter than the Air-1 case. Heating rates also followed a similar trend. Figure 6 shows the heat flux normalized by the square root of the static pressure. In this figure one sees that the trends are uniform, however, an unexplained difficultly is why the measured normalized heating rate was about 2 times the value predicted by the Fay-Riddell correlation (shown as the straight line). The Fay-Riddell prediction was also checked through the NATA code which showed it value to be slightly higher than calculated by the catalysis program primarily because of higher predicted shock properties of temperature and pressure. It should also be noted that the NATA code calculated a Lewis number of 1.27 while the catalysis code calculated a Lewis number of less than 1 at around 0.9 (dependent on the experimental point). Lewis numbers less than 1 result when the total atomic mass fraction is high. This signifies that the ratio of diffusion transport to the surface is less then thermal transport away from the surface.

Figure 7 shows the recombination coefficient determined for nitrogen. The curve is relatively flat representing a process that has low activation energy (gas phase recombination reactions also have low activation energies). The magnitude of the results compare favorably with those reported by Scott (1981) (his values ranged from 0.014 to 0.02 in the similar temperature range). The results also compare with Kolodziej and Stewart (1987) in which they reported values from 0.008 to 0.016. Both papers reported relatively low activation energies which also compare with these results.

Figure 8 shows the recombination coefficient determined for oxygen. These curves show the difficultly of data repeatability between runs. Further, one would expect the results for the 14% O_2 to fall between Air-2 and the 11% O_2 if the Air-1 data could be thrown out for an obvious reason. Attempts to explain the day to day discrepancies included verification of voltage measurement, current measurement and water flow rates. The voltage measurement check involved connecting a Simpson meter across the supply lines. The Simpson meter measurement checked with the value measured by the Norbitrol. Current calibrations were done several times with no major discrepancies found. Finally, the water turbine meter position was changed so that no sudden flow disturbance occurred upstream. No flow change was observed for this correction. In a review of arc jet data, the only major discrepancies was with the oxygen supply pressure and actual flowrate. However, the product of these two were the same for both air runs as it should be. Further, the column pressures were very similar indicating that the arc jet column conditions were identical. So what explanations remain to be explored? Surface change is the most likely candidate. Perhaps emissivity changes or perhaps nitrogen changes. Another explanation may be thermocouple relocation, however, two independent nitrogen runs were repeatable. Another possibility may be the model location was not repeatable and this should be explored further. In summary, further testing is required.

TABLE 1

SUMMARY OF CONDITIONS STUDIED FOR CATALYTIC RECOMBINATION ON RCG COATED HRSI MATERIAL

| Run Number	Gases	Flowrate kg/sec	Current amps	Enthalpy MJ/kg	Calorimeter W/cm^2	Surface Temp. K
2-749-1	Air	0.0368	350	7.31	50.66	1448.6
2-749-2	Air	0.0368	376	7.75	56.06	1467.5
2-749-3	Air	0.0368	403	8.20	60.08	1497.3
2-749-4	Air	0.0368	426	8.66	63.72	1518.0
2-749-5	Air	0.0368	450	9.11	64.42	1539.8
2-749-6	Air	0.0368	451	9.14	64.95	1540.1
2-749-7	Air	0.0368	425	8.71	60.12	1527.1
2-749-8	Air	0.0368	401	8.25	56.59	1500.5
2-749-9	Air	0.0368	376	7.80	52.53	1485.8
2-749-10	Air	0.0368	349	7.34	47.57	1458.0
2-749-11	N_2	0.0363	402	7.89	42.20	1349.8
2-751-1	N_2	0.0381	400	7.65	53.74	1410.7
2-751-2	N_2	0.0381	451	8.46	60.07	1460.0
2-751-3	N_2	0.0381	501	9.13	61.81	1490.2
2-751-4	N_2	0.0381	524	9.42	66.88	1506.0
2-751-5	N_2	0.0381	550	9.83	68.21	1523.9
2-751-6	N_2	0.0390	500	9.04	64.44	1493.2
2-751-7	N_2	0.0368	400	7.74	50.08	1411.6
2-751-8	14% O_2	0.0558	301	5.25	46.95	1468.0
2-751-9	14% O_2	0.0563	326	5.64	49.86	1508.7
2-751-10	14% O_2	0.0563	341	5.84	53.48	1529.2
2-753-1	Air	0.0359	351	7.43	61.18	1502.3
2-753-2	Air	0.0359	374	7.80	66.56	1527.1
2-753-3	Air	0.0359	401	8.30	69.21	1557.0
2-753-4	11% O_2	0.0368	300	6.45	51.96	1455.9
2-753-5	11% O_2	0.0368	326	6.97	52.94	1479.1
2-753-6	11% O_2	0.0368	351	7.35	58.24	1502.1
2-753-7	11% O_2	0.0368	376	7.79	60.89	1521.9
2-753-8	11% O_2	0.0395	400	7.77	65.07	1552.2
2-753-9	11% O_2	0.0372	326	6.89	51.54	1485.9
2-753-11	N_2	0.0368	400	7.69	50.92	1402.2
2-753-12	N_2	0.0368	501	9.32	67.50	1495.7
2-753-13	N_2	0.0368	550	10.05	71.48	1536.1
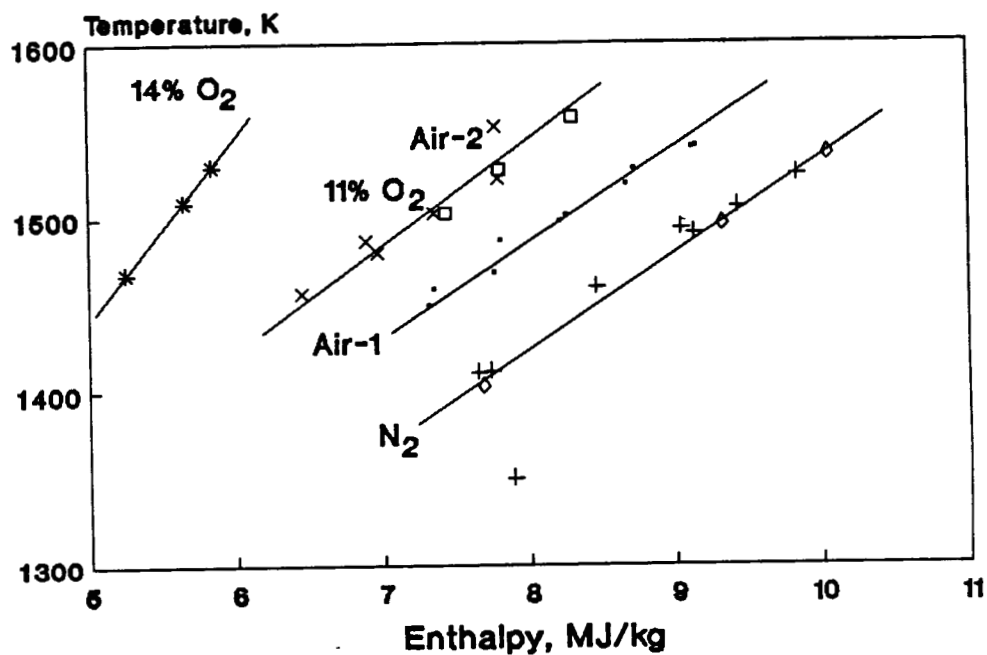
Figure 5. Surface Temperatures Measured at the Stagnation Point of an HRSI RCG Coated Hemisphere as a Function of Enthalpy for all Test Conditions.
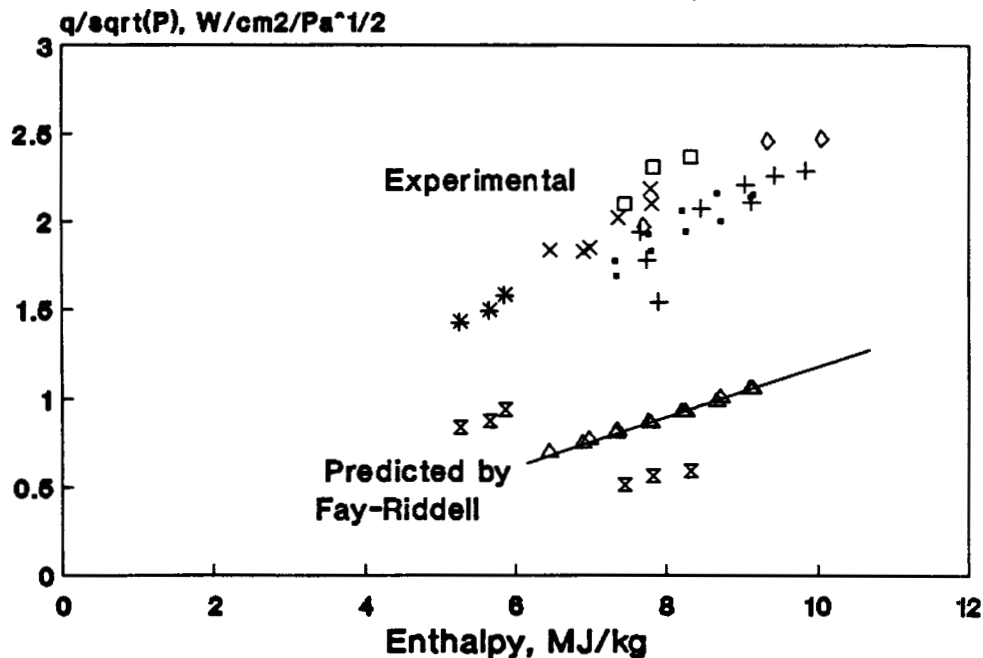

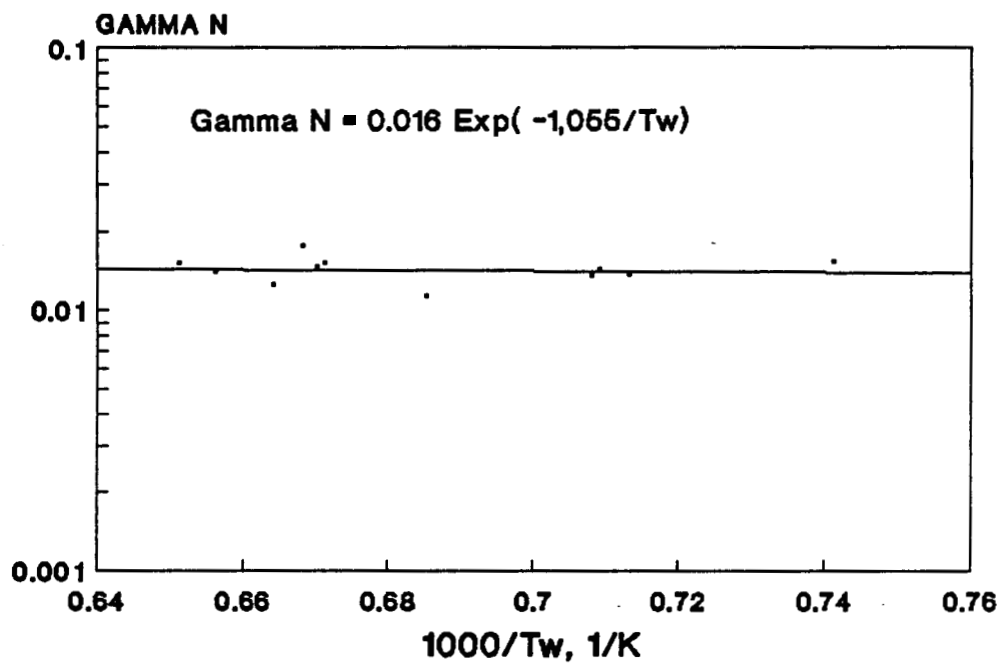Figure 6. Normalized Heat Flux Measured as a Function of Total Enthalpy.

GAMMA N

Gamma N = 0.016 Exp(-1,055/Tw)

1000/Tw, 1/K

Figure 7. Nitrogen Recombination Coefficient as a Function of Temperature – Arrhenius Plot.

GAMMA O
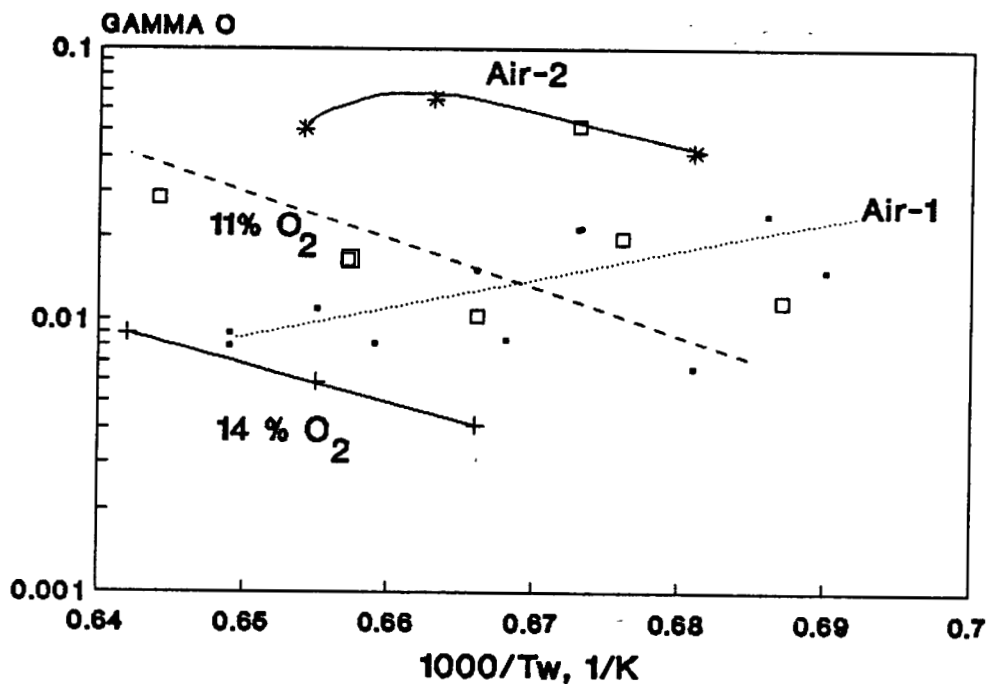
Air-2

11% O$_2$

Air-1

14 % O$_2$

1000/Tw, 1/K

Figure 8. Oxygen Recombination Coefficient as a Function of Temperature – for Various Run Conditions.

After the test program was completed, the surface of the slug calorimeter was analyzed by energy dispersive X-Ray spectroscopy (EDAX). The results showed that the silver coating agglomerated on the surface exposing the copper surface underneath. Further elemental analysis showed that the small clumps of were primarily composed of silver with a trace of oxygen and copper. The valley areas were primarily composed of copper with some oxygen, however, the oxygen level was not high enough to represent cuprous oxide or cupric oxide on the surface.

CONCLUSIONS

1. HRSI - RCG coated thermal protection materials are relatively inert to nitrogen atom recombination.

2. Repeatability within runs was possible, however, outside of runs, results were not as repeatable.

3. The scatter of data observed between runs for oxygen recombination experiments cannot be attributed to differences in voltage, current, or water flowrate measurement errors.

3. The silver coated calorimeters changed surface structure morphology during the test program with silver agglomerates forming on the surface.

RECOMMENDATIONS

1. Continue experimental measurements. Temperature ranges between 1600 and 1800 K should be investigated.

2. Attempt catalytic measurements on RCC (Reinforced Carbon Carbon) materials.

3. Determine why the surface temperature differed by about 60 K for identical arc heater conditions between the two air runs.

4. Include emissivity measurements in all catalysis test programs.

5. Insure that all arc jet measurements are reliable before testing.

REFERENCES

AMX-100 MUX Gas Analyzer Instruction Manual, UTI P/N 12719, UTI Inc. Sunnyvale California

Bade, W.L.; Yos, J.M.: The Nata Code, Vols. 1 & 2 NASA CR-141743, 1976.

Fay, J.A.; Riddell, F.R.: J. Aeronaut. Sci. Vol 25, 1958, pp 73-83.

Goulard, R.: On Catalytic Recombination Rates in Hypersonic Stagnation Heat Transfer," Jet Propulsion, Vol 28, 1958, pp 737-745.

Kolodziej,P.; Stewart, D.A.: Nitrogen Recombination on High Temperature Reusable Surface Insulation and the Analysis of its Effect on Surface Catalysis. AIAA-87-1637, June 1987.

McLafferty, F.W.: Interpretation of Mass Spectra, W.A. Benjamin Inc., 1973.

Rochelle,W.C.; Battley, H.H.;Grimaud, J.E.;Tillian, D.J.; Murray, L.P.; Lueke, W.J.; and Heaton, T.M.: Orbiter TPS Development and Certification Testing at the NASA/JSC 10 MW Atmospheric Reentry Materials and Structures Evaluation Facility. AIAA-83-0147, Jan 1983.

Scott,C.D.: Catalytic Recombination of Nitrogen and Oxygen on High-Temperature Reusable Surface Insulation. Aerothermodynamics and Planetary Entry, edited by A.L.Crosbie, Vol.77 of Progress in Astronautics and Aeronautics, 1981, pp 192-212.

Scott, C.D.; Derry, S.M.: Catalytic Recombination and The Space Shuttle Heating, AIAA-82-0841, 1982.

Scott, C.D.: Catalytic Recombination of Nitrogen and Oxygen on Iron-Cobalt-Chromia Spinel, AIAA-83-0585, 1983.

Stewart, D.A.; Rakich, J.V.; Lanfranco, M.J.: Catalytic Surface Effects Experiment on the Space Shuttle, AIAA-81-1143, 1981.

Willey, R.J.: The Identification of Excited Species in Arc Jet Flow. NASA CR-171990, 1983.

REPORT DOCUMENTATION PAGE

1. Report No. NASA CR 172118	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle NASA/ASEE Summer Faculty Fellowship Program--1988 Volume 2	5. Report Date February 1989
	6. Performing Organization Code

7. Author(s) Richard B. Bannerot and Stanley H. Goldstein, Editors	8. Performing Organization Report No.

9. Performing Organization Name and Address The University of Houston--University Park Houston, Texas 77004	10. Work Unit No.
	11. Contract or Grant No. NGT-44-005-803

12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546	13. Type of Report and Period Covered Contractor Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The 1988 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began in 1965 at JSC and in 1964 nationally, are (1) to further the professional knowledge of qualified engineering and science faculty members; (2) to stimulate an exchange of ideas between participants and NASA; (3) to enrich and refresh the research and teaching activities of participants' institutions; and (4) to contribute to the research objectives of the NASA Centers.

Each faculty fellow spent 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1988. Volume 1 contains reports 1 through 14, and volume 2 contains reports 15 through 26.

17. Key Words (Suggested by Author(s))	18. Distribution Statement Unclassified-Unlimited

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of pages 371	22. Price NTIS

For sale by the National Technical Information Service, Springfield, VA 22161-2171

JSC Form 1424 (Rev Jan 88) (Ethernet Jan 88)